

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 22, 2004, 12:01:03 ; Search time 111 Seconds  
(without alignments)  
1476.374 Million cell updates/sec

Title: US-10-069-541-6  
Perfect score: 2972  
Sequence: 1 MAFHVEGLIAIIVFYLLILL.....EAFLDVDSSPEGSGTEDNLQ 580

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2972	100.0	580	4	AAB74665	Aab74665 Human hig
2	2972	100.0	580	4	AAB86837	Aab86837 Human CHO
3	2972	100.0	580	6	ABU08979	Abu08979 Human hig
4	2972	100.0	580	7	ADD50649	Add50649 High-affi
5	2972	100.0	580	7	ADD50639	Add50639 Human hig
6	2972	100.0	580	7	ADD50648	Add50648 High-affi
7	2972	100.0	580	7	ADD50647	Add50647 High-affi
8	2820	94.9	580	4	AAB74664	Aab74664 Rat high
9	2820	94.9	580	7	ADD50643	Add50643 Rat high-

10	2810	94.5	580	4	AA72388	Aay72388 Mouse P4P
11	2801	94.2	580	4	AAB74666	Aab74666 Mouse hig
12	2795	94.0	580	7	ADD50641	Add50641 Mouse hig
13	2795	94.0	580	7	ADD50661	Add50661 Mouse hig
14	1453	48.9	576	4	AAB74663	Aab74663 C. elegan
15	1453	48.9	576	7	ADD50645	Add50645 C. elegan
16	723.5	24.3	278	4	ABB71601	Abb71601 Drosophil
17	344	11.6	492	4	AAB96805	Aab96805 Putative
18	317.5	10.7	675	5	ABP69833	Abp69833 Human pol
19	311.5	10.5	675	4	AAB85102	Aab85102 Novel hum
20	311.5	10.5	675	6	ABJ37936	Abj37936 NOVX prot
21	310.5	10.4	675	4	AAB60093	Aab60093 Human tra
22	310.5	10.4	675	5	ABB80599	Abb80599 Human sbg
23	308.5	10.4	662	2	AAR73595	Aar73595 Cotranspo
24	306	10.3	660	4	ABG30151	Abg30151 Novel hum
25	299.5	10.1	546	5	ABP65617	Abp65617 Bifidobac
26	298.5	10.0	675	5	AAU77134	Aau77134 Human sod
27	298	10.0	672	2	AAR73593	Aar73593 Cotranspo
28	297.5	10.0	471	6	ABU24801	Abu24801 Protein e
29	295	9.9	596	6	ABB99498	Abb99498 Amino aci
30	295	9.9	655	7	ADC97745	Adc97745 Na/glucos
31	294	9.9	670	5	ABP69135	Abp69135 Human pol
32	287	9.7	659	2	AA731221	Aay31221 Human SAA
33	286	9.6	678	5	ABU05343	Abu05343 Pancreas-
34	283	9.5	484	4	AAB96255	Aab96255 Putative
35	279	9.4	454	6	ABU25621	Abu25621 Protein e
36	277.5	9.3	596	4	AAE06614	Aae06614 Human pro
37	277.5	9.3	596	4	AAE08088	Aae08088 Human tra
38	277.5	9.3	596	5	ABG31594	Abg31594 Human tra
39	277.5	9.3	596	5	ABB80588	Abb80588 Human sbg
40	277.5	9.3	596	5	AAO14199	Aao14199 Human tra
41	277.5	9.3	596	6	ABU10299	Abu10299 Novel hum
42	277.5	9.3	643	6	ABB99497	Abb99497 Amino aci
43	277	9.3	596	6	ABU10300	Abu10300 Rabbit so
44	273	9.2	648	5	ABP69832	Abp69832 Human pol
45	273	9.2	681	5	ABU05344	Abu05344 Pancreas-

# ALIGNMENTS

```

RESULT 1
AAB74665
ID   AAB74665 standard; protein; 580 AA.
XX
AC   AAB74665;
XX
DT   01-JUN-2001   (first entry)
XX
DE   Human high affinity choline transporter protein.
XX
KW   High affinity choline transporter; cho-1; Alzheimer's disease; diagnosis.
XX
OS   Homo sapiens.
XX
PN   WO200116315-A1.
XX

```

PD 08-MAR-2001.  
 XX  
 PF 18-AUG-2000; 2000WO-JP005545.  
 XX  
 PR 27-AUG-1999; 99JP-00240642.  
 PR 27-DEC-1999; 99JP-00368991.  
 XX  
 PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.  
 XX  
 PI Haga T, Okuda T;  
 XX  
 DR WPI; 2001-226688/23.  
 DR N-PSDB; AAF81712.  
 XX  
 PT New rat and human spinal cord high affinity choline transporters, useful  
 PT in diagnosis of Alzheimer's disease and screening promoters as drugs for  
 PT treating Alzheimer's disease.  
 XX  
 PS Claim 8; Page 76-78; 90pp; Japanese.  
 XX  
 CC The present sequence represents a human (Homo sapiens) high affinity  
 CC choline transporter protein designated cho-1. The cho-1 protein has  
 CC nootropic and neuroprotective activities. The cho-1 polynucleotide and  
 CC protein can be used for the diagnosis of diseases related to the  
 CC expression of cho-1 by comparing the cho-1 polynucleotide sequence in a  
 CC sample to that of a control. Drug compositions containing the cho-1  
 CC protein or expression promoters or inhibitors of cho-1 are useful for  
 CC treating disorders characterised by abnormal levels of cho-1, such as  
 CC Alzheimer's disease  
 XX  
 SQ Sequence 580 AA;

Query Match 100.0%; Score 2972; DB 4; Length 580;  
 Best Local Similarity 100.0%; Pred. No. 3.6e-290;  
 Matches 580; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTA	60
Db	1	MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTA	60
Qy	61	TWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ	120
Db	61	TWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ	120
Qy	121	IYGKRMGLLFIPALMGEMFWAAAFSALGATISVIIDVDMHISVIISALIATLYTLVGG	180
Db	121	IYGKRMGLLFIPALMGEMFWAAAFSALGATISVIIDVDMHISVIISALIATLYTLVGG	180
Qy	181	LYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSSEVYSW	240
Db	181	LYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSSEVYSW	240
Qy	241	LDSFLLMLGGIPWQAYFQRVLSSSSATYAQVLSFLAAFGLVMAIPAILIGAIGASTDW	300
Db	241	LDSFLLMLGGIPWQAYFQRVLSSSSATYAQVLSFLAAFGLVMAIPAILIGAIGASTDW	300
Qy	301	NQTAYGLPDPKTTTEADMILPIVLQYLCVPYISFFGLGAVSAAVMSSADSSILSASSMFA	360

```

Db      301  NQTAYGLPDPKTTTEADMILPIVLQYLCVPYISFFGLGAVSAAVMSSADSSILSASSMFA 360
QY      361  RNIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQ 420
Db      361  RNIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQ 420
QY      421  LLCVLFVKGTNTYGAVAGYVSGFLRLITGGEPYLYLQPLIFYPGYPPDDNGIYNQKFPFK 480
Db      421  LLCVLFVKGTNTYGAVAGYVSGFLRLITGGEPYLYLQPLIFYPGYPPDDNGIYNQKFPFK 480
QY      481  TLAMVTSFLTNICISYLAKEYLFESGTLPPKLDVFDVAVARHSEENMDKTILVKENIKLD 540
Db      481  TLAMVTSFLTNICISYLAKEYLFESGTLPPKLDVFDVAVARHSEENMDKTILVKENIKLD 540
QY      541  ELALVKPRQSMTLSSTFTNKEAFLDVDSSPEGSGTEDNLQ 580
Db      541  ELALVKPRQSMTLSSTFTNKEAFLDVDSSPEGSGTEDNLQ 580

```

RESULT 2

AAB86837

ID AAB86837 standard; protein; 580 AA.

XX

AC AAB86837;

XX

DT 26-NOV-2001 (first entry)

XX

DE Human CHOT protein.

XX

KW CHOT; human; choline transporter; chromosome 2q11-13; nootropic;  
 KW neuroprotective; gene therapy; antisense therapy; degenerative disease;  
 KW cognitive disorder; Alzheimer's disease.

XX

OS Homo sapiens.

XX

PN DE10009055-A1.

XX

PD 30-AUG-2001.

XX

PF 28-FEB-2000; 2000DE-01009055.

XX

PR 28-FEB-2000; 2000DE-01009055.

XX

PA (BRUE/) BRUESS M.

PA (BOEN/) BOENISCH H.

XX

PI Bruess M, Boenisch H;

XX

DR WPI; 2001-590709/67.

DR N-PSDB; AAH49207.

XX

PT A new gene encoding human choline transporter, designated hCHOT is  
 PT located on chromosome 2q11-13 and is useful to treat degenerative  
 PT disorders such as Alzheimer's disease.

XX

PS Disclosure; Page 11; 12pp; German.



XX  
CC This invention describes a novel gene encoding human choline transporter,  
CC designated hCHOT which is located on chromosome 2q11-13. The products of  
CC the invention have nootropic and neuroprotective activity and can be used  
CC for gene or antisense therapy. (I) is used to treat degenerative disease,  
CC particularly cognitive disorders such as Alzheimer's disease. Sense and  
CC antisense oligonucleotides derived from the gene may be used in  
CC diagnostics and other techniques. This sequence represents the human CHOT  
CC protein described in the invention  
XX  
SQ Sequence 580 AA;

Query Match 100.0%; Score 2972; DB 4; Length 580;  
Best Local Similarity 100.0%; Pred. No. 3.6e-290;  
Matches 580; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTA 60  
|  
Db 1 MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTA 60  
  
Qy 61 TWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ 120  
|  
Db 61 TWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ 120  
  
Qy 121 IYGKRMGGLLFIPALMGEMFWAAAI FSALGATISVIIDVDMHISVIISALIATLYTLVGG 180  
|  
Db 121 IYGKRMGGLLFIPALMGEMFWAAAI FSALGATISVIIDVDMHISVIISALIATLYTLVGG 180  
  
Qy 181 LYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSSEVYSW 240  
|  
Db 181 LYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSSEVYSW 240  
  
Qy 241 LDSFLLMLGGIPWQAYFQRLSSSSATYAQVLSFLAAFGLVMAIPAILIGAIGASTDW 300  
|  
Db 241 LDSFLLMLGGIPWQAYFQRLSSSSATYAQVLSFLAAFGLVMAIPAILIGAIGASTDW 300  
  
Qy 301 NQTAYGLPDPKTTTEADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFA 360  
|  
Db 301 NQTAYGLPDPKTTTEADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFA 360  
  
Qy 361 RNIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQ 420  
|  
Db 361 RNIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQ 420  
  
Qy 421 LLCVLFVKGTNTYGAVAGYVSGFLRITGGEPYLYLQPLIFYPGYPPDDNGIYNQKFPFK 480  
|  
Db 421 LLCVLFVKGTNTYGAVAGYVSGFLRITGGEPYLYLQPLIFYPGYPPDDNGIYNQKFPFK 480  
  
Qy 481 TLAMVTSFLTNICISYLA KYLFESGTLPPKLDVFDVAVVARHSEENMDKTILVKKENIKLD 540  
|  
Db 481 TLAMVTSFLTNICISYLA KYLFESGTLPPKLDVFDVAVVARHSEENMDKTILVKKENIKLD 540  
  
Qy 541 ELALVKPRQSMTLSSTFTNKEAFLDVDSSPEGSGTEDNLQ 580  
|  
Db 541 ELALVKPRQSMTLSSTFTNKEAFLDVDSSPEGSGTEDNLQ 580

RESULT 3

ABU08979

ID ABU08979 standard; protein; 580 AA.

XX

AC ABU08979;

XX

DT 13-JUN-2003 (first entry)

XX

DE Human high affinity choline transporter, HACT.

XX

KW Human; HACT; high affinity choline transporter; pain;

KW neurotransmitter biosynthesis; learning and memory; aging; epilepsy;

KW neurological disorder; spasticity; myoclonus; muscle spasm;

KW muscle hyperactivity; stroke; head trauma; neuronal cell death;

KW multiple sclerosis; spinal chord injury; dystonia; Alzheimer's disease;

KW Myasthenia Gravis; multi-infarct dementia; AIDS dementia;

KW Parkinson's disease; Huntington's disease; amyotrophic lateral sclerosis;

KW ALS; attention deficit disorder; organic brain syndrome; schizophrenia;

KW nicotine addiction; memory disorder; cognitive disorder.

XX

OS Homo sapiens.

XX

PN US6500643-B1.

XX

PD 31-DEC-2002.

XX

PF 07-SEP-2000; 2000US-00657252.

XX

PR 07-SEP-2000; 2000US-00657252.

XX

PA (UYFL ) UNIV FLORIDA.

XX

PI Wu D, Gu Y, Millard WJ, He Y;

XX

DR WPI; 2003-361535/34.

DR N-PSDB; ABX94338.

XX

PT Novel isolated polynucleotide (I) that encodes high affinity choline

PT transporter protein, useful for preventing, treating or ameliorating

PT neurological and cognitive disorders such as Alzheimer's or Parkinson's

PT disease.

XX

PS Claim 1; Col 21-24; 20pp; English.

XX

CC The invention relates to an isolated polynucleotide which encodes a high

CC affinity choline transporter (HACT) protein appearing as ABU08979. Also

CC included are a polynucleotide encoding a fragment consisting of at least

CC about 50 amino acids of the HACT protein, a vector comprising the

CC polynucleotide, a composition comprising a vector comprising a

CC polynucleotide which comprises at least about 12 contiguous nucleic acids

CC of a polynucleotide appearing as ABX94339 (encoding choline

CC acetyltransferase), a recombinant host cell which comprises the vector

CC (used to express the HACT protein or fragment). The polynucleotide is

CC useful as a probe or primer to detect the presence of HACT polynucleotide

CC in a sample, such as a biological sample, or for screening for test

CC agents which bind to the polynucleotide. A pharmaceutical composition

CC comprising the polynucleotide is useful for preventing, treating or

CC ameliorating neurological and cognitive disorders e.g. pain, spasticity,  
CC myoclonus, muscle spasm, muscle hyperactivity, epilepsy, stroke, head  
CC trauma, neuronal cell death, multiple sclerosis, spinal chord injury,  
CC dystonia, Alzheimer's disease, myasthenia gravis, multi- infarct  
CC dementia, AIDS dementia, Parkinson's disease, Huntington's disease,  
CC amyotrophic lateral sclerosis (ALS), attention deficit disorder, nicotine  
CC addiction, organic brain syndromes, schizophrenia or memory and cognitive  
CC disorders. HACT is thought to be the rate limiting step in cholinergic  
CC neurotransmitter biosynthesis and regeneration (cholinergic transmissions  
CC are crucial to brain functions such as learning and memory). The present  
CC sequence represents human HACT

XX

SQ Sequence 580 AA;

Query Match 100.0%; Score 2972; DB 6; Length 580;  
Best Local Similarity 100.0%; Pred. No. 3.6e-290;  
Matches 580; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTA 60
      |||
Db      1 MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTA 60

Qy     61 TWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ 120
      |||
Db     61 TWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ 120

Qy    121 IYKRMGGLLFIPALMGEMFWAAAI FSALGATISVIIDVDMHISVIISALIATLYTLVGG 180
      |||
Db    121 IYKRMGGLLFIPALMGEMFWAAAI FSALGATISVIIDVDMHISVIISALIATLYTLVGG 180

Qy    181 LYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSSEVYSW 240
      |||
Db    181 LYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSSEVYSW 240

Qy    241 LDSFLLMLGGIPWQAYFQRLSSSSATYAQVLSFLAAGCLVMAIPAILIGAIGASTDW 300
      |||
Db    241 LDSFLLMLGGIPWQAYFQRLSSSSATYAQVLSFLAAGCLVMAIPAILIGAIGASTDW 300

Qy    301 NQTAYGLPDPKTTEEADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFA 360
      |||
Db    301 NQTAYGLPDPKTTEEADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFA 360

Qy    361 RNIIQLSFRQNASDKEIVWMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQ 420
      |||
Db    361 RNIIQLSFRQNASDKEIVWMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQ 420

Qy    421 LLCVLFVKGTNTYGAVAGYVSGFLRITGGEPYLYLQPLIFYPGYYPDDNGIYNQKFPFK 480
      |||
Db    421 LLCVLFVKGTNTYGAVAGYVSGFLRITGGEPYLYLQPLIFYPGYYPDDNGIYNQKFPFK 480

Qy    481 TLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDVAVVARHSEENMDKTILVKENIKLD 540
      |||
Db    481 TLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDVAVVARHSEENMDKTILVKENIKLD 540

Qy    541 ELALVKPRQSMTLSSTFTNKEAFLDVDSSPEGSGTEDNLQ 580
      |||
Db    541 ELALVKPRQSMTLSSTFTNKEAFLDVDSSPEGSGTEDNLQ 580
```

RESULT 4

ADD50649

ID ADD50649 standard; protein; 580 AA.

XX

AC ADD50649;

XX

DT 15-JAN-2004 (first entry)

XX

DE High-affinity choline transporter (CHT) associated protein sequence #3.

XX

KW High-affinity choline transporter; CHT; cholinergic function;

KW Parkinson's disease; Huntington's disease; Alzheimer's disease;

KW schizophrenia; dysautonomia; myasthenia gravis; brain;

KW cholinergic signalling; antiparkinsonian; anticonvulsant; nootropic;

KW neuroprotective; neuroleptic.

XX

OS Unidentified.

XX

PN US2003114399-A1.

XX

PD 19-JUN-2003.

XX

PF 23-JUL-2001; 2001US-00911077.

XX

PR 23-JUL-2001; 2001US-00911077.

XX

PA (BLAK/) BLAKELY R D.

PA (APPA/) APPARSUNDARAM S.

PA (FERG/) FERGUSON S.

XX

PI Blakely RD, Apparsundaram S, Ferguson S;

XX

DR WPI; 2003-810914/76.

XX

PT Novel isolated polynucleotide encoding human or mouse high affinity  
PT choline transporter polypeptide, useful in gene therapy to increase  
PT cholinergic function in a cell of a patient suffering from Alzheimer's  
PT disease.

XX

PS Disclosure; SEQ ID NO 12; 74pp; English.

XX

CC The present invention relates to the isolation of polynucleotide  
CC sequences encoding human and mouse high-affinity choline transporter  
CC (hCHT and mCHT respectively), and the proteins they encode. The gene  
CC encoding hCHT is located on chromosome 2q12. The polynucleotide sequence  
CC encoding hCHT, is useful for expressing hCHT recombinantly. The hCHT  
CC polynucleotide sequence when delivered to a cell, increases cholinergic  
CC function in the cell that is in a patient having Parkinson's disease,  
CC Huntington's disease, Alzheimer's disease, schizophrenia, dysautonomia or  
CC myasthenia gravis. The hCHT antibody is useful for controlling  
CC transporter CHT proteins to the brain, and for treating the above  
CC mentioned diseases. The antibody is also useful for diagnosing the above  
CC mentioned disorders and to detect the influence of cholinergic  
CC signalling. The present protein sequence of unknown function is provided  
CC in the electronic sequence data but is not mentioned in the printed

CC specification. Note: The sequence data for this patent was obtained in  
CC electronic format directly from the USPTO web site at seqdata.uspto.gov.  
XX  
SQ Sequence 580 AA;

Query Match 100.0%; Score 2972; DB 7; Length 580;  
Best Local Similarity 100.0%; Pred. No. 3.6e-290;  
Matches 580; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTA 60
        |
Db      1 MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTA 60

Qy     61 TWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ 120
        |
Db     61 TWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ 120

Qy    121 IYGKRMGGLLFIPALMGEMFWAAAFSALGATISVIIDVDMHISVIISALIATLYTLVGG 180
        |
Db    121 IYGKRMGGLLFIPALMGEMFWAAAFSALGATISVIIDVDMHISVIISALIATLYTLVGG 180

Qy    181 LYSVAYTDVVQLFCIFVGLWISVPPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSSEVYSW 240
        |
Db    181 LYSVAYTDVVQLFCIFVGLWISVPPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSSEVYSW 240

Qy    241 LDSFLLMLGGIPWQAYFQRLVSSSSATYAQVLSFLAAGCLVMAIPAILIGAIGASTDW 300
        |
Db    241 LDSFLLMLGGIPWQAYFQRLVSSSSATYAQVLSFLAAGCLVMAIPAILIGAIGASTDW 300

Qy    301 NQTAYGLPDPKTTEEADMILPIVLQYLCVPYISFFGLGAVSAAVMSSADSSILSASSMFA 360
        |
Db    301 NQTAYGLPDPKTTEEADMILPIVLQYLCVPYISFFGLGAVSAAVMSSADSSILSASSMFA 360

Qy    361 RNIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQ 420
        |
Db    361 RNIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQ 420

Qy    421 LLCVLFVKGTNTYGAVAGYVSGFLRITGGEPYLYLQPLIFYPGYPPDDNGIYNQKFPEK 480
        |
Db    421 LLCVLFVKGTNTYGAVAGYVSGFLRITGGEPYLYLQPLIFYPGYPPDDNGIYNQKFPEK 480

Qy    481 TLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDVAVARHSEENMDKTILVKENIKLD 540
        |
Db    481 TLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDVAVARHSEENMDKTILVKENIKLD 540

Qy    541 ELALVKPRQSMTLSSSTFTNKEAFLDVDSSPEGSGTEDNLQ 580
        |
Db    541 ELALVKPRQSMTLSSSTFTNKEAFLDVDSSPEGSGTEDNLQ 580
```

RESULT 5

ADD50639

ID ADD50639 standard; protein; 580 AA.

XX

AC ADD50639;

XX

DT 15-JAN-2004 (first entry)

XX  
 DE Human high-affinity choline transporter (hCHT).  
 XX  
 KW Human; high-affinity choline transporter; hCHT; cholinergic function;  
 KW Parkinson's disease; Huntington's disease; Alzheimer's disease;  
 KW schizophrenia; dysautonomia; myasthenia gravis; brain;  
 KW cholinergic signalling; antiparkinsonian; anticonvulsant; nootropic;  
 KW neuroprotective; neuroleptic.  
 XX  
 OS Homo sapiens.  
 XX  
 PN US2003114399-A1.  
 XX  
 PD 19-JUN-2003.  
 XX  
 PF 23-JUL-2001; 2001US-00911077.  
 XX  
 PR 23-JUL-2001; 2001US-00911077.  
 XX  
 PA (BLAK/) BLAKELY R D.  
 PA (APPA/) APPARSUNDARAM S.  
 PA (FERG/) FERGUSON S.  
 XX  
 PI Blakely RD, Apparsundaram S, Ferguson S;  
 XX  
 DR WPI; 2003-810914/76.  
 DR N-PSDB; ADD50638.  
 XX  
 PT Novel isolated polynucleotide encoding human or mouse high affinity  
 PT choline transporter polypeptide, useful in gene therapy to increase  
 PT cholinergic function in a cell of a patient suffering from Alzheimer's  
 PT disease.  
 XX  
 PS Claim 1; SEQ ID NO 2; 74pp; English.  
 XX  
 CC The present invention relates to the isolation of polynucleotide  
 CC sequences encoding human and mouse high-affinity choline transporter  
 CC (hCHT and mCHT respectively), and the proteins they encode. The gene  
 CC encoding hCHT is located on chromosome 2q12. The polynucleotide sequence  
 CC encoding hCHT, is useful for expressing hCHT recombinantly. The hCHT  
 CC polynucleotide sequence when delivered to a cell, increases cholinergic  
 CC function in the cell that is in a patient having Parkinson's disease,  
 CC Huntington's disease, Alzheimer's disease, schizophrenia, dysautonomia or  
 CC myasthenia gravis. The hCHT antibody is useful for controlling  
 CC transporter CHT proteins to the brain, and for treating the above  
 CC mentioned diseases. The antibody is also useful for diagnosing the above  
 CC mentioned disorders and to detect the influence of cholinergic  
 CC signalling. The present sequence represents hCHT. Note: The sequence data  
 CC for this patent was obtained in electronic format directly from the USPTO  
 CC web site at seqdata.uspto.gov.  
 XX  
 SQ Sequence 580 AA;

Query Match 100.0%; Score 2972; DB 7; Length 580;  
 Best Local Similarity 100.0%; Pred. No. 3.6e-290;  
 Matches 580; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTA	60
Db	1	MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTA	60
Qy	61	TWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ	120
Db	61	TWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ	120
Qy	121	IYGKRMGGLLFIPALMGEMFWAAAFISALGATISVIIDVDMHISVIISALIATLYTLVGG	180
Db	121	IYGKRMGGLLFIPALMGEMFWAAAFISALGATISVIIDVDMHISVIISALIATLYTLVGG	180
Qy	181	LYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSEVYSW	240
Db	181	LYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSEVYSW	240
Qy	241	LDSFLLMLGGIPWQAYFQRVLSSSSATYAQVLSFLAAFGLVMAIPAILIGAIGASTDW	300
Db	241	LDSFLLMLGGIPWQAYFQRVLSSSSATYAQVLSFLAAFGLVMAIPAILIGAIGASTDW	300
Qy	301	NQTAYGLPDPKTTTEADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFA	360
Db	301	NQTAYGLPDPKTTTEADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFA	360
Qy	361	RNIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQ	420
Db	361	RNIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQ	420
Qy	421	LLCVLFVKGTNTYGAVAGYVSGFLFRITGGEPYLYLQPLIFYPGYYPDDNGIYNQKFPEK	480
Db	421	LLCVLFVKGTNTYGAVAGYVSGFLFRITGGEPYLYLQPLIFYPGYYPDDNGIYNQKFPEK	480
Qy	481	TLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDVAVVARHSEENMDKTILVKKENIKLD	540
Db	481	TLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDVAVVARHSEENMDKTILVKKENIKLD	540
Qy	541	ELALVKPRQSMTLSSTFTNKEAFLDVDSSPEGSGTEDNLQ	580
Db	541	ELALVKPRQSMTLSSTFTNKEAFLDVDSSPEGSGTEDNLQ	580

RESULT 6

ADD50648

ID ADD50648 standard; protein; 580 AA.

XX

AC ADD50648;

XX

DT 15-JAN-2004 (first entry)

XX

DE High-affinity choline transporter (CHT) associated protein sequence #2.

XX

KW High-affinity choline transporter; CHT; cholinergic function;

KW Parkinson's disease; Huntington's disease; Alzheimer's disease;

KW schizophrenia; dysautonomia; myasthenia gravis; brain;

KW cholinergic signalling; antiparkinsonian; anticonvulsant; nootropic;

KW neuroprotective; neuroleptic.

XX

OS Unidentified.  
XX  
PN US2003114399-A1.  
XX  
PD 19-JUN-2003.  
XX  
PF 23-JUL-2001; 2001US-00911077.  
XX  
PR 23-JUL-2001; 2001US-00911077.  
XX  
PA (BLAK/) BLAKELY R D.  
PA (APPA/) APPARSUNDARAM S.  
PA (FERG/) FERGUSON S.  
XX  
PI Blakely RD, Apparsundaram S, Ferguson S;  
XX  
DR WPI; 2003-810914/76.  
XX  
PT Novel isolated polynucleotide encoding human or mouse high affinity  
PT choline transporter polypeptide, useful in gene therapy to increase  
PT cholinergic function in a cell of a patient suffering from Alzheimer's  
PT disease.  
XX  
PS Disclosure; SEQ ID NO 11; 74pp; English.  
XX  
CC The present invention relates to the isolation of polynucleotide  
CC sequences encoding human and mouse high-affinity choline transporter  
CC (hCHT and mCHT respectively), and the proteins they encode. The gene  
CC encoding hCHT is located on chromosome 2q12. The polynucleotide sequence  
CC encoding hCHT, is useful for expressing hCHT recombinantly. The hCHT  
CC polynucleotide sequence when delivered to a cell, increases cholinergic  
CC function in the cell that is in a patient having Parkinson's disease,  
CC Huntington's disease, Alzheimer's disease, schizophrenia, dysautonomia or  
CC myasthenia gravis. The hCHT antibody is useful for controlling  
CC transporter CHT proteins to the brain, and for treating the above  
CC mentioned diseases. The antibody is also useful for diagnosing the above  
CC mentioned disorders and to detect the influence of cholinergic  
CC signalling. The present protein sequence of unknown function is provided  
CC in the electronic sequence data but is not mentioned in the printed  
CC specification. Note: The sequence data for this patent was obtained in  
CC electronic format directly from the USPTO web site at seqdata.uspto.gov.  
XX  
SQ Sequence 580 AA;

Query Match 100.0%; Score 2972; DB 7; Length 580;  
Best Local Similarity 100.0%; Pred. No. 3.6e-290;  
Matches 580; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTA 60  
|  
Db 1 MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTA 60  
QY 61 TWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ 120  
|  
Db 61 TWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ 120  
QY 121 IYGKRMGGLLFIPALMGEMFWAAAFSALGATISVIIDVDMHISVIISALIATLYTLVGG 180



Db	121	 IYGKRMGGLLFIPALMGEMFWAAAFSALGATISVIIDVDMHISVIISALIATLYTLVGG	180
Qy	181	LYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSSEVYSW	240
Db	181	 LYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSSEVYSW	240
Qy	241	LDSFLLMLGGIPWQAYFQRVLSSSSATYAQVLSFLAAFGCLVMAIPAILIGAIGASTDW	300
Db	241	 LDSFLLMLGGIPWQAYFQRVLSSSSATYAQVLSFLAAFGCLVMAIPAILIGAIGASTDW	300
Qy	301	NQTAYGLPDPKTTEEADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFA	360
Db	301	 NQTAYGLPDPKTTEEADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFA	360
Qy	361	RNIYQLSFRQNASDKEIVWMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQ	420
Db	361	 RNIYQLSFRQNASDKEIVWMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQ	420
Qy	421	LLCVLFVKGTNTYGAVAGYVSGFLRITGGEPYLYLQPLIFYPGYPPDDNGIYNQKFPFK	480
Db	421	 LLCVLFVKGTNTYGAVAGYVSGFLRITGGEPYLYLQPLIFYPGYPPDDNGIYNQKFPFK	480
Qy	481	TLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDVAVVARHSEENMDKTILVKENIKLD	540
Db	481	 TLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDVAVVARHSEENMDKTILVKENIKLD	540
Qy	541	ELALVKPRQSMTLSSTFTNKEAFLDVDSSPEGSGTEDNLQ	580
Db	541	 ELALVKPRQSMTLSSTFTNKEAFLDVDSSPEGSGTEDNLQ	580

RESULT 7

ADD50647

ID ADD50647 standard; protein; 580 AA.

XX

AC ADD50647;

XX

DT 15-JAN-2004 (first entry)

XX

DE High-affinity choline transporter (CHT) associated protein sequence #1.

XX

KW High-affinity choline transporter; CHT; cholinergic function;

KW Parkinson's disease; Huntington's disease; Alzheimer's disease;

KW schizophrenia; dysautonomia; myasthenia gravis; brain;

KW cholinergic signalling; antiparkinsonian; anticonvulsant; nootropic;

KW neuroprotective; neuroleptic.

XX

OS Unidentified.

XX

PN US2003114399-A1.

XX

PD 19-JUN-2003.

XX

PF 23-JUL-2001; 2001US-00911077.

XX

PR 23-JUL-2001; 2001US-00911077.

XX  
PA (BLAK/) BLAKELY R D.  
PA (APPA/) APPARSUNDARAM S.  
PA (FERG/) FERGUSON S.  
XX  
PI Blakely RD, Apparsundaram S, Ferguson S;  
XX  
DR WPI; 2003-810914/76.  
XX  
PT Novel isolated polynucleotide encoding human or mouse high affinity  
PT choline transporter polypeptide, useful in gene therapy to increase  
PT cholinergic function in a cell of a patient suffering from Alzheimer's  
PT disease.  
XX  
PS Disclosure; SEQ ID NO 10; 74pp; English.  
XX  
CC The present invention relates to the isolation of polynucleotide  
CC sequences encoding human and mouse high-affinity choline transporter  
CC (hCHT and mCHT respectively), and the proteins they encode. The gene  
CC encoding hCHT is located on chromosome 2q12. The polynucleotide sequence  
CC encoding hCHT, is useful for expressing hCHT recombinantly. The hCHT  
CC polynucleotide sequence when delivered to a cell, increases cholinergic  
CC function in the cell that is in a patient having Parkinson's disease,  
CC Huntington's disease, Alzheimer's disease, schizophrenia, dysautonomia or  
CC myasthenia gravis. The hCHT antibody is useful for controlling  
CC transporter CHT proteins to the brain, and for treating the above  
CC mentioned diseases. The antibody is also useful for diagnosing the above  
CC mentioned disorders and to detect the influence of cholinergic  
CC signalling. The present protein sequence of unknown function is provided  
CC in the electronic sequence data but is not mentioned in the printed  
CC specification. Note: The sequence data for this patent was obtained in  
CC electronic format directly from the USPTO web site at seqdata.uspto.gov.  
XX  
SQ Sequence 580 AA;

Query Match 100.0%; Score 2972; DB 7; Length 580;  
Best Local Similarity 100.0%; Pred. No. 3.6e-290;  
Matches 580; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTA	60
Db	1	MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTA	60
Qy	61	TWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ	120
Db	61	TWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ	120
Qy	121	IYGKRMGGLLFIPALMGEMFWAAAFSALGATISVIIDVDMHISVIISALIATLYTLVGG	180
Db	121	IYGKRMGGLLFIPALMGEMFWAAAFSALGATISVIIDVDMHISVIISALIATLYTLVGG	180
Qy	181	LYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSEVYSW	240
Db	181	LYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSEVYSW	240
Qy	241	LDSFLLMLGGIPWQAYFQRVLSSSSATYAQVLSFLAAGCLVMAIPAILIGAIGASTDW	300

Db 241 LDSFLLMLGGIPWQAYFQRVLSSSSATYAQVLSFLAAGCLVMAIPAILIGAIGASTDW 300  
 Qy 301 NQTAYGLPDPKTTEEADMILPIVLQYLCVYISFFGLGAVSAAVMSSADSSILSASSMFA 360  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 301 NQTAYGLPDPKTTEEADMILPIVLQYLCVYISFFGLGAVSAAVMSSADSSILSASSMFA 360  
 Qy 361 RNIIQLSFRQNASDKEIVVMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQ 420  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 361 RNIIQLSFRQNASDKEIVVMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQ 420  
 Qy 421 LLCVLFVKGTNTYGAVAGYVSGFLRITGGEPYLYLQPLIFYPGYPPDDNGIYNQKFPFK 480  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 421 LLCVLFVKGTNTYGAVAGYVSGFLRITGGEPYLYLQPLIFYPGYPPDDNGIYNQKFPFK 480  
 Qy 481 TLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDVAVVARHSEENMDKTILVKENIKLD 540  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 481 TLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDVAVVARHSEENMDKTILVKENIKLD 540  
 Qy 541 ELALVKPRQSMTLSSTFTNKEAFLDVDSSPEGSGTEDNLQ 580  
 ||||||||||||||||||||||||||||||||||||||||||||  
 Db 541 ELALVKPRQSMTLSSTFTNKEAFLDVDSSPEGSGTEDNLQ 580

RESULT 8

AAB74664

ID AAB74664 standard; protein; 580 AA.

XX

AC AAB74664;

XX

DT 01-JUN-2001 (first entry)

XX

DE Rat high affinity choline transporter protein.

XX

KW High affinity choline transporter; cho-1; Alzheimer's disease; diagnosis.

XX

OS Rattus norvegicus.

XX

PN WO200116315-A1.

XX

PD 08-MAR-2001.

XX

PF 18-AUG-2000; 2000WO-JP005545.

XX

PR 27-AUG-1999; 99JP-00240642.

PR 27-DEC-1999; 99JP-00368991.

XX

PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.

XX

PI Haga T, Okuda T;

XX

DR WPI; 2001-226688/23.

DR N-PSDB; AAF81711.

XX

PT New rat and human spinal cord high affinity choline transporters, useful  
 PT in diagnosis of Alzheimer's disease and screening promoters as drugs for  
 PT treating Alzheimer's disease.

XX

PS Claim 5; Page 69-71; 90pp; Japanese.

XX

CC The present sequence represents a rat (*Rattus norvegicus*) high affinity  
CC choline transporter protein designated cho-1. The cho-1 protein has  
CC nootropic and neuroprotective activities. The cho-1 polynucleotide and  
CC protein can be used for the diagnosis of diseases related to the  
CC expression of cho-1 by comparing the cho-1 polynucleotide sequence in a  
CC sample to that of a control. Drug compositions containing the cho-1  
CC protein or expression promoters or inhibitors of cho-1 are useful for  
CC treating disorders characterised by abnormal levels of cho-1, such as  
CC Alzheimer's disease

XX

SQ Sequence 580 AA;

Query Match 94.9%; Score 2820; DB 4; Length 580;  
Best Local Similarity 93.1%; Pred. No. 7.6e-275;  
Matches 540; Conservative 24; Mismatches 16; Indels 0; Gaps 0;

```
Qy      1 MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGFTMTA 60
      | |||||:||||:||||| |||||:||||:||||| ||||| |||||
Db      1 MPFHVEGLVAIILFYLLIFLVGIWAAWKTKNSGNAEERSEAIIVGGRDIGLLVGFTMTA 60

Qy     61 TWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ 120
      ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db     61 TWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ 120

Qy    121 IYGKRMGGLLFIPALMGEMFWAAAI FSALGATISVIIDVDMHISVIISALIATLYTLVGG 180
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db    121 IYGKRMGGLLFIPALMGEMFWAAAI FSALGATISVIIDVDNISVIVSALIAILYTLVGG 180

Qy    181 LYSVAYTDVVQLFCIFVGLWISVFPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSSEVYSW 240
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db    181 LYSVAYTDVVQLFCIFVGLWISVFPFALSHPAVTDIGFTAVHAKYQSPWLGTIESVEVYTW 240

Qy    241 LDSFLLMLGGIPWQAYFQRLSSSSATYAQVLSFLAAGCLVMAIPAILIGAIGASTDW 300
      ||:||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db    241 LDNFLLMLGGIPWQAYFQRLSSSSATYAQVLSFLAAGCLVMALPAICIGAIGASTDW 300

Qy    301 NQTAYGLPDPKTTEEADMILPIVLQYLCVPYISFFGLGAVSAAVMSSADSSILSASSMFA 360
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db    301 NQTAYGFDPDKTKEEADMILPIVLQYLCVPYISFFGLGAVSAAVMSSADSSILSASSMFA 360

Qy    361 RNIYQLSFRQNASDKEIVWMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQ 420
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db    361 RNIYQLSFRQNASDKEIVWMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIIIFPQ 420

Qy    421 LLCVLFVKGTNTYGAVAGYVSGFLRITGGEPYLYLQPLIFYPGYYPDDNGIYNQKFPFK 480
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db    421 LLCVLFVKGTNTYGAVAGYIFGLFLRITGGEPYLYLQPLIFYPGYYPDKNGIYNQKFPFK 480

Qy    481 TLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDVAVVARHSEENMDKTILVKNENIKLD 540
      ||:|||| ||||:||||| ||||| ||||| ||||| ||||| |||||
Db    481 TLSMVTSFFTNICVSYLAKYLFESGTLPPKLDIFDVAVSRHSEENMDKTILVRNENIKLN 540

Qy    541 ELALVKPRQSM TLSSTFTNKEAFLDVDSSPEGSGTEDNLQ 580
      ||| |||||:||||| ||||| ||||| |||||
Db    541 ELAPVKPRQSL TLSSTFTNKEALLDVDSSPEGSGTEDNLQ 580
```

RESULT 9

ADD50643

ID ADD50643 standard; protein; 580 AA.

XX

AC ADD50643;

XX

DT 15-JAN-2004 (first entry)

XX

DE Rat high-affinity choline transporter (rCHT).

XX

KW Rat; high-affinity choline transporter; rCHT; cholinergic function;

KW Parkinson's disease; Huntington's disease; Alzheimer's disease;

KW schizophrenia; dysautonomia; myasthenia gravis; brain;

KW cholinergic signalling; antiparkinsonian; anticonvulsant; nootropic;

KW neuroprotective; neuroleptic.

XX

OS Rattus sp.

XX

PN US2003114399-A1.

XX

PD 19-JUN-2003.

XX

PF 23-JUL-2001; 2001US-00911077.

XX

PR 23-JUL-2001; 2001US-00911077.

XX

PA (BLAK/) BLAKELY R D.

PA (APPA/) APPARSUNDARAM S.

PA (FERG/) FERGUSON S.

XX

PI Blakely RD, Apparsundaram S, Ferguson S;

XX

DR WPI; 2003-810914/76.

DR N-PSDB; ADD50642.

XX

PT Novel isolated polynucleotide encoding human or mouse high affinity

PT choline transporter polypeptide, useful in gene therapy to increase

PT cholinergic function in a cell of a patient suffering from Alzheimer's

PT disease.

XX

PS Example 1; SEQ ID NO 6; 74pp; English.

XX

CC The present invention relates to the isolation of polynucleotide

CC sequences encoding human and mouse high-affinity choline transporter

CC (hCHT and mCHT respectively), and the proteins they encode. The gene

CC encoding hCHT is located on chromosome 2q12. The polynucleotide sequence

CC encoding hCHT, is useful for expressing hCHT recombinantly. The hCHT

CC polynucleotide sequence when delivered to a cell, increases cholinergic

CC function in the cell that is in a patient having Parkinson's disease,

CC Huntington's disease, Alzheimer's disease, schizophrenia, dysautonomia or

CC myasthenia gravis. The hCHT antibody is useful for controlling

CC transporter CHT proteins to the brain, and for treating the above

CC mentioned diseases. The antibody is also useful for diagnosing the above

CC mentioned disorders and to detect the influence of cholinergic

CC signalling. The present sequence represents rat CHT (rCHT). Note: The

CC sequence data for this patent was obtained in electronic format directly  
CC from the USPTO web site at seqdata.uspto.gov.  
XX  
SQ Sequence 580 AA;

Query Match 94.9%; Score 2820; DB 7; Length 580;  
Best Local Similarity 93.1%; Pred. No. 7.6e-275;  
Matches 540; Conservative 24; Mismatches 16; Indels 0; Gaps 0;

```
Qy      1 MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTA 60
        | | | | | : | | : | | | | | | | | : | | | | : | | | | | | | | | | | | | |
Db      1 MPFHVEGLVAIILFYLLIFLVGIWAAWTKNSGNAEERSEAIIVGGRDIGLLVGGFTMTA 60

Qy     61 TWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ 120
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     61 TWVGGGYINGTAEAVYGPCCGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ 120

Qy    121 IYGKRMGGLLFIPALMGEMFWAAAFSALGATISVIIDVDMHISVIISALIATLYTLVGG 180
        | | | | | | | | | | | | | | | | | | | | | | | | : | | | : | | | | | | | |
Db    121 IYGKRMGGLLFIPALMGEMFWAAAFSALGATISVIIDVDVNISVIVSALIAILYTLVGG 180

Qy    181 LYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSSEVYSW 240
        | | | | | | | | | | : | | | | | | | | | | | | | | | | | | | | | | | |
Db    181 LYSVAYTDVVQLFCIFIGLWISVPFALSHPAVTDIGFTAVHAKYQSPWLGTIESVEVYTW 240

Qy    241 LDSFLLLMLGGIPWQAYFQRLSSSSATYAQVLSFLAAGCLVMAIPAILIGAIGASTDW 300
        | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    241 LDNFLLLLMLGGIPWQAYFQRLSSSSATYAQVLSFLAAGCLVMALPAICIGAIGASTDW 300

Qy    301 NQTAYGLPDPKTTEEADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFA 360
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    301 NQTAYGFDPDKTKEEADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFA 360

Qy    361 RNIYQLSFRQNASDKEIVWMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIIVFPQ 420
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    361 RNIYQLSFRQNASDKEIVWMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIIIFPQ 420

Qy    421 LLCVLFVKGTNTYGAAGYVSGFLRITGGEPYLYLQPLIFYPGYYPDDNGIYNQKFPFK 480
        | | | | | : | | | | | | | | : | | | | | | | | | | | | | | | | | | | | |
Db    421 LLCVLFVKGTNTYGAAGYIFGLRITGGEPYLYLQPLIFYPGYYPDKNGIYNQKFPFK 480

Qy    481 TLMVTSFSLTNICISYLAKYLFESGTLPPKLDVFDVAVRHSEENMDKTILVKNENIKLD 540
        | | : | | | | | | | : | | | | | | | | | | | | | | | | | | | | | | : | | |
Db    481 TLMVTSFSLTNICISYLAKYLFESGTLPPKLDIFDAVSRHSEENMDKTILVRNENIKLN 540

Qy    541 ELALVKPRQSMSTLSSTFTNKEAFLDVDSSPEGSGTEDNLQ 580
        | | | | | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    541 ELAPVKPRQSLTSLSTFTNKEALLDVDSSPEGSGTEDNLQ 580
```

RESULT 10  
AAY72388  
ID AAY72388 standard; protein; 580 AA.  
XX  
AC AAY72388;  
XX  
DT 24-APR-2001 (first entry)

```

XX      Mouse P4P6B1 OMA (obese mice adipocyte) protein.
DE
XX
KW      Mouse; OMA protein; obese mice adipocyte; P4P6B1;
KW      fuel metabolism disorder; therapy; obesity; diabetes; gene therapy;
KW      anorectic; antidiabetic.
XX
OS      Mus sp.
XX
PN      WO200078950-A2.
XX
PD      28-DEC-2000.
XX
PF      13-JUN-2000; 2000WO-US016217.
XX
PR      22-JUN-1999; 99US-0141515P.
XX
PA      (AMYL-) AMYLIN PHARM INC.
XX
PI      Sierzega M, Albrandt K;
XX
DR      WPI; 2001-112322/12.
DR      N-PSDB; AAD02457.
XX
PT      Novel obese mice adipocyte polypeptides useful in diagnosis and treatment
PT      of disorders of fuel metabolism such as obesity or diabetes.
XX
PS      Claim 11; Fig 3; 83pp; English.
XX
CC      The present sequence is mouse OMA (obese mice adipocyte) protein encoded
CC      by P4P6B1 cDNA. The P4P6B1 cDNA fragment was generated by RNA
CC      fingerprinting using random primers P4 and P6. OMA is used as a
CC      diagnostic reagent for diagnosing a disorder of fuel metabolism in an
CC      underweight or an overweight individual, by detecting the transcription
CC      level of a gene encoding OMA, which is induced or repressed in an
CC      individual by a factor such as genetic obesity, fasting and refeeding of
CC      a fasted individual. OMA is useful in the generation of antibodies, for
CC      use in pharmaceutical compositions and for studying DNA/protein
CC      interactions. Nucleic acids encoding OMA are involved in gene therapy. An
CC      inhibitor of OMA or an antisense oligonucleotide that inhibits expression
CC      of OMA are useful for treating disorders of fuel metabolism such as
CC      obesity or diabetes
XX
SQ      Sequence 580 AA;

```

```
Qy      1 MAFHVEGLIAIIVFYLLILLVGIWAAWRKTKNSGSAEERSEAIIVGGRDIGLLVGGMTMTA 60
        |:|||||:|||||:||||: ||||| ||||| ||||| ||||| ||||| |||||
Db      1 MSFHVEGLVAIILFYLLILLVGIWAAWKTKNSGNPEERSEAIIVGGRDIGLLVGGMTMTA 60

Qy     61 TWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGVTMLDPFQQ 120
        | ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db     61 TWVGGGYINGTAEAVYGPGLAWAQAPIGYSLSLILGGLFFAKPMRSKGVTMLDPFQQ 120
```

Qy	121	IYGKRMGGLLFIPALMGEMFWAAAFSALGATISVIIDVDMHISVIISALIATLYTLVGG	180
		:     :	
Db	121	IYGKRMGGLLFIPALMGEMFWAAAFSALGATISVIIDVDVNISVIVSALIAILYTLVGG	180
Qy	181	LYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSEVYSW	240
		:     :	
Db	181	LYSVAYTDVVQLFCIFIGLWISVPFALSHPAVTDIGFTAVHAKYQSPWLGTIESVEVYTW	240
Qy	241	LDSFLLMLGGIPWQAYFQRLSSSSATYAQVLSFLAAGCLVMAIPAILIGAIGASTDW	300
		:     :	
Db	241	LDNFLLMLGGIPWQAYFQRLSSSSATYAQVLSFLAAGCLVMALPAICIGAIGASTDW	300
Qy	301	NQTAYGLPDPKTEEADMILPIVLQYLCPVYISFFGLGAVSAAMSSADSSILSASSMFA	360
Db	301	NQTAYGYDPDKTKEEADMILPIVLQYLCPVYISFFGLGAVSAAMSSADSSILSASSMFA	360
Qy	361	RNIYQLSFRQNASDKEIVVMRITVVFVGASATAMALLTKTVYGLWYLSSDLVYIVIFPQ	420
		:	
Db	361	RNIYQLSFRQNASDKEIVVMRITVLVFGASATAMALLTKTVYGLWYLSSDLVYIIIFPQ	420
Qy	421	LLCVLFVKGTNTYGAVAGYVSGLEFLRITGGEPYLYLQPLIFYPGYPPDDNGIYNQKFPEK	480
		:     :	
Db	421	LLCVLFIKGTNTYGAVAGYIFGLEFLRITGGEPYLYLQPLIFYPGYSDKNGIYNQRFPEK	480
Qy	481	TLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDVAVVARHSEENMDKTILVKENIKLD	540
		:     :     :	
Db	481	TLSMVTSEFTNICVSYLAKYLFESGTLPPKLDVFDVAVVARHSEENMDKTILVRNENIKLN	540
Qy	541	ELALVKPRQSMTLSSTFTNKEAFLDVDSSPEGSGTEDNLQ	580
Db	541	ELAPVKPRQSLTSLSTFTNKEALLDVDSSPEGSGTEDNLQ	580

RESULT 11

AAB74666

ID AAB74666 standard; protein; 580 AA.

XX

AC AAB74666;

XX

DT 01-JUN-2001 (first entry)

XX

DE Mouse high affinity choline transporter protein.

XX

KW High affinity choline transporter; cho-1; Alzheimer's disease; diagnosis.

XX

OS Mus musculus.

XX

PN WO200116315-A1.

XX

PD 08-MAR-2001.

XX

PF 18-AUG-2000; 2000WO-JP005545.

XX

PR 27-AUG-1999; 99JP-00240642.

PR 27-DEC-1999; 99JP-00368991.

XX

PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.



XX  
 PI Haga T, Okuda T;  
 XX  
 DR WPI; 2001-226688/23.  
 DR N-PSDB; AAF81713.  
 XX  
 PT New rat and human spinal cord high affinity choline transporters, useful  
 PT in diagnosis of Alzheimer's disease and screening promoters as drugs for  
 PT treating Alzheimer's disease.  
 XX  
 PS Claim 11; Page 82-85; 90pp; Japanese.  
 XX  
 CC The present sequence represents a mouse (*Mus musculus*) high affinity  
 CC choline transporter protein designated cho-1. The cho-1 protein has  
 CC nootropic and neuroprotective activities. The cho-1 polynucleotide and  
 CC protein can be used for the diagnosis of diseases related to the  
 CC expression of cho-1 by comparing the cho-1 polynucleotide sequence in a  
 CC sample to that of a control. Drug compositions containing the cho-1  
 CC protein or expression promoters or inhibitors of cho-1 are useful for  
 CC treating disorders characterised by abnormal levels of cho-1, such as  
 CC Alzheimer's disease  
 XX  
 SQ Sequence 580 AA;

Query Match 94.2%; Score 2801; DB 4; Length 580;  
 Best Local Similarity 92.8%; Pred. No. 6.3e-273;  
 Matches 538; Conservative 23; Mismatches 19; Indels 0; Gaps 0;

Qy	1	MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFMTA	60
		:     :    :            :    :	
Db	1	MSFHVEGLVAIILFYLLIFLVGIWAAWTKNSGNPEEHSEAIIVGGRDIGLLVGGFMTA	60
Qy	61	TWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ	120
Db	61	TWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ	120
Qy	121	IYGKRMGGLLFIPALMGEMFWAAAFSALGATISVIIDVDMHISVIISALIATLYTLVGG	180
		:    :	
Db	121	IYGKRMGGLLFIPALMGEMFWAAAFSALGATISVIIDVDVNISVIVSALIAILYTLVGG	180
Qy	181	LYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSSEVYSW	240
		:    :            :     :	
Db	181	LYSVAYTDVVQLFCIFIGLWISVPFALSHPAVTDIGFTAVHAKYQSPWLGTIESVEVYTW	240
Qy	241	LDSFLLMLGGIPWQAYFQRVLSSSSATYAQVLSFLAAGCLVMAIPAILIGAIGASTDW	300
		:     :    :	
Db	241	LDNFLLMLGGIPWQAYFQRVLSSSSATYAQVLSFLAAGCLVMALPAICIGAIGASTDW	300
Qy	301	NQTAYGLPDPKTTTEADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFA	360
Db	301	NQTAYGYPDPKTKKEADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFA	360
Qy	361	RNIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQ	420
		:	
Db	361	RNIYQLSFRQNASDKEIVWVMRITVLVFGASATAMALLTKTVYGLWYLSSDLVYIIIFPQ	420
Qy	421	LLCVLFVKGTNTYGAVAGYVSGFLFRITGGEPYLYLQPLIFYPGYPPDDNGIYNQKFPFK	480





Db

541 ELAPVKPRQSLTSLSTFTNKEALLDVDSSPEGSGTEDNLQ 580

RESULT 13

ADD50661

ID ADD50661 standard; protein; 580 AA.

XX

AC ADD50661;

XX

DT 15-JAN-2004 (first entry)

XX

DE Mouse high-affinity choline transporter (mCHT) #2.

XX

KW Mouse; high-affinity choline transporter; mCHT; cholinergic function;

KW Parkinson's disease; Huntington's disease; Alzheimer's disease;

KW schizophrenia; dysautonomia; myasthenia gravis; brain;

KW cholinergic signalling; antiparkinsonian; anticonvulsant; nootropic;

KW neuroprotective; neuroleptic.

XX

OS Mus sp.

XX

PN US2003114399-A1.

XX

PD 19-JUN-2003.

XX

PF 23-JUL-2001; 2001US-00911077.

XX

PR 23-JUL-2001; 2001US-00911077.

XX

PA (BLAK/) BLAKELY R D.

PA (APPA/) APPARSUNDARAM S.

PA (FERG/) FERGUSON S.

XX

PI Blakely RD, Apparsundaram S, Ferguson S;

XX

DR WPI; 2003-810914/76.

DR N-PSDB; ADD50660.

XX

PT Novel isolated polynucleotide encoding human or mouse high affinity

PT choline transporter polypeptide, useful in gene therapy to increase

PT cholinergic function in a cell of a patient suffering from Alzheimer's

PT disease.

XX

PS Disclosure; SEQ ID NO 24; 74pp; English.

XX

CC The present invention relates to the isolation of polynucleotide

CC sequences encoding human and mouse high-affinity choline transporter

CC (hCHT and mCHT respectively), and the proteins they encode. The gene

CC encoding hCHT is located on chromosome 2q12. The polynucleotide sequence

CC encoding hCHT, is useful for expressing hCHT recombinantly. The hCHT

CC polynucleotide sequence when delivered to a cell, increases cholinergic

CC function in the cell that is in a patient having Parkinson's disease,

CC Huntington's disease, Alzheimer's disease, schizophrenia, dysautonomia or

CC myasthenia gravis. The hCHT antibody is useful for controlling

CC transporter CHT proteins to the brain, and for treating the above

CC mentioned diseases. The antibody is also useful for diagnosing the above

CC mentioned disorders and to detect the influence of cholinergic

CC signalling. The present sequence represents mCHT. Note: The sequence data  
CC for this patent was obtained in electronic format directly from the USPTO  
CC web site at seqdata.uspto.gov.

XX

SQ Sequence 580 AA;

Query Match 94.0%; Score 2795; DB 7; Length 580;  
Best Local Similarity 92.6%; Pred. No. 2.5e-272;  
Matches 537; Conservative 23; Mismatches 20; Indels 0; Gaps 0;

```
Qy      1 MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTA 60
      | | | | | : | | : | | | | | | | | | | | : | | | | : | | | | | | | | | | | |
Db      1 MPFHVEGLVAIILFYLLIFLVGIWAAWKTKNNGNPEERSEAIIVGGRDIGLLVGGFTMTA 60

Qy     61 TWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ 120
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     61 TWVGGGYINGTAEAVYGP GCLAWAHAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFKQ 120

Qy    121 IYGKRMGGLLFIPALMGEMFWAAAI FSALGATISVIIDVDMHISVIISAL IATLYTLVGG 180
      | | | | | | | | | | | | | | | | | | | | | | | | | | | : | | | : | | | | |
Db    121 IYGKRMGGLLFIPALMGEMFWAAAI FSALGATISVIIDVDVNISVIVSAL IAILYTLVGG 180

Qy    181 LYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSSEVYSW 240
      | | | | | | | | | | | | | | | | | | | | | | | | | | | : | | | : | | : |
Db    181 LYSVAYTDVVQLFCIFIGLWISVPFALSHPAVTDIGFTAVHAKYQSPWLGTIESVEVYTW 240

Qy    241 LDSFLLMLGGIPWQAYFQ RVLSSSSATYAQVLSFLAAFGLVMAIPAILIGAIGASTDW 300
      | | : | | | | | | | | | | | | | | | | | | | | | | | | | : | | | | | | |
Db    241 LDNFLMLGGIPWQAYFQ RVLSSSSATYAQVLSFLAAFGLVMALPAICIGAIGASTDW 300

Qy    301 NQTAYGLPDPKTTEEADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFA 360
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    301 NQTAYGYDPDKTKEEADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFA 360

Qy    361 RNIYQLSFRQNASDKEIVWVMRITVFVGASATAMALLTKTVYGLWYLSDDLVIYIIFPQ 420
      | | | | | | | | | | | | | | | | | | | | | | | | | | | : | | | |
Db    361 RNIYQLSFRQNASDKEIVWVMRITVLVFGASATAMALLTKTVYGLWYLSDDLVIYIIFPQ 420

Qy    421 LLCVL FVKGTNTYGA VAGYVSGFLRITGGEPYLYLQPLIFYPGYYPDDNGIYNQKFPFK 480
      | | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | : | | |
Db    421 LLCVLFIKGTNTYGA VAGYIFGLFLRITGGEPYLYLQPLIFYPGYSDKNGIYNQRFPFK 480

Qy    481 TLAMVTSFLTNICISYLA KYLFESGTLPPKLDVFD AVVARHSEENMDKTILVKNENIKLD 540
      | | : | | | | | | | : | | | | | | | | | | | | | | | | | | | | | : | | | |
Db    481 TLSMVT SFFTNICVSYLA KYLFESGTLPPKLDVFD AVVARHSEENMDKTILVRNENIKLN 540

Qy    541 ELALVKPRQSMTLSSTFTNKEAFLDVDSSPEGSGTEDNLQ 580
      | | | | | | | | | | | | | | | | | | | | | | | |
Db    541 ELAPVKPRQSLTLSSTFTNKEALLDVDSSPEGSGTEDNLQ 580
```

RESULT 14

AAB74663

ID AAB74663 standard; protein; 576 AA.

XX

AC AAB74663;

XX

DT 01-JUN-2001 (first entry)  
 XX  
 DE C. elegans high affinity choline transporter protein.  
 XX  
 KW High affinity choline transporter; cho-1; Alzheimer's disease; diagnosis.  
 XX  
 OS Caenorhabditis elegans.  
 XX  
 PN WO200116315-A1.  
 XX  
 PD 08-MAR-2001.  
 XX  
 PF 18-AUG-2000; 2000WO-JP005545.  
 XX  
 PR 27-AUG-1999; 99JP-00240642.  
 PR 27-DEC-1999; 99JP-00368991.  
 XX  
 PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.  
 XX  
 PI Haga T, Okuda T;  
 XX  
 DR WPI; 2001-226688/23.  
 DR N-PSDB; AAF81710.  
 XX  
 PT New rat and human spinal cord high affinity choline transporters, useful  
 PT in diagnosis of Alzheimer's disease and screening promoters as drugs for  
 PT treating Alzheimer's disease.  
 XX  
 PS Claim 2; Page 62-64; 90pp; Japanese.  
 XX  
 CC The present sequence represents a Caenorhabditis elegans high affinity  
 CC choline transporter protein designated cho-1. The cho-1 protein has  
 CC nootropic and neuroprotective activities. The cho-1 polynucleotide and  
 CC protein can be used for the diagnosis of diseases related to the  
 CC expression of cho-1 by comparing the cho-1 polynucleotide sequence in a  
 CC sample to that of a control. Drug compositions containing the cho-1  
 CC protein or expression promoters or inhibitors of cho-1 are useful for  
 CC treating disorders characterised by abnormal levels of cho-1, such as  
 CC Alzheimer's disease  
 XX  
 SQ Sequence 576 AA;

Query Match 48.9%; Score 1453; DB 4; Length 576;  
 Best Local Similarity 50.5%; Pred. No. 4.7e-137;  
 Matches 295; Conservative 95; Mismatches 150; Indels 44; Gaps 9;

Qy 7 GLIAIIVFYLLILLVGIWAAWRITKNSGSAEER----SEAIIVGGRDIGLLVGGFTMTATW 62  
 |::||: ||:||||:|||| |::| | :| :: ||:| || | |||||  
 Db 6 GIVAIVFFYVLILVVGIWAGRKSKSSKELESEAGAATEEVMLAGRNIGTLVGIFTMTATW 65  
 Qy 63 VGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQIY 122  
 ||| |||||::| || |||:|::|::|::| || || :|:||||| |  
 Db 66 VGGAYINGTAEALY--NGGLIGCQAPVGYAISLVMGGLLFAKKMREEGYITMLDPFQHKY 123  
 Qy 123 GKRMGGLLFIPALMGEMFWAAAFSALGATISVIIDVDMHISVIISALIATLYTLVGGLY 182  
 |:|:|::|::|::| || ||| |||||:|::|::|: || :| || || || |  
 Db 124 GQRIGGLMYVPALLGETFWTAAILSALGATLSVILGIDMNASVTLSACIAVFYFTFTGGYY 183

QY 183' SVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDS-SEVYSWL 241  
 :|||||:|||||: || |: | || | :| : | :  
 Db 184 AVAYTDVVQLFCIFVGLWVCVPAAMVHDGAKDISRNA-----GDWIGEIGGFKETSLWI 237

QY 242 DSFLLMLGGIPWQAYFQRVLSSSSATYAQVLSFLAAGCIVMAIPAILIGAIGASTDWN 301  
 | |||: ||||| ||||| :| || |||: ||::||| |||| :|||  
 Db 238 DCMLLLVFGGIPWQVYFQRVLSKTAHGAQTLSEFVAGVGCILMAIPPALIGAIARNTDWR 297

QY 302 QTAYGLPDPKTTEEA-----DMILPIVLQYLCVPYISFFGLGAVSAAVMSSADSSILSA 355  
 | | : | |: :||:| || | ::| |||||:||||  
 Db 298 MTDYSPWNNGTKVESIPDPKRNMVVPLVFQYLTPRWVAFI GLGAVSAAVMSSADSSVLSA 357

QY 356 SSMFARNIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYI 415  
 :||| ||::| :||:|: ||| : | || ||| ::||| | :|||:  
 Db 358 ASMFAHNIWKLTI RPHASEKEVIIVMRIAICVGIMATIMALTIQSIYGLWYLCADLVYV 417

QY 416 VIFPQLLCVLFVKGTNTYGA VAGYVSGFLFRITGGEPYLYLQPLIFYPGYYPDDNGIYNQ 475  
 ::|||:|:| :|||:| || ||: ||| : | || | :| : |  
 Db 418 ILFPQLLCVVMRPSNTYGS LAGYAVGLVLR LIGGEPLVSLPAFFHYPMY---TDGV--Q 472

QY 476 KFPFKTLAMVTSFLT NICISYLA KYLFESGTLPPKLDVFDVAV---ARHSEENMDKTILV 532  
 |||:| ||::| | :| :: ||:| | |: || || | : |  
 Db 473 YFPFRTTAMLSSMATIYIVSIQSEKLFKSGRLSPEWDVMGCVVNIPIDHVLPLSDVSEAV 532

QY 533 KNENIKL-----DELALVKPRQSMTLSSTFTN 559  
 :| : : || |: | : || :|  
 Db 533 SSETLNMKAPNGTPAPVHPNQPSDENTLLHPYSDQSYYSTNSN 576

# RESULT 15

ADD50645

ID ADD50645 standard; protein; 576 AA.

XX

AC ADD50645;

XX

DT 15-JAN-2004 (first entry)

XX

DE C. elegans CHOI protein.

XX

KW High-affinity choline transporter; CHT; cholinergic function;

KW Parkinson's disease; Huntington's disease; Alzheimer's disease;

KW schizophrenia; dysautonomia; myasthenia gravis; brain;

KW cholinergic signalling; antiparkinsonian; anticonvulsant; nootropic;

KW neuroprotective; neuroleptic; CHOI.

XX

OS Caenorhabditis elegans.

XX

PN US2003114399-A1.

XX

PD 19-JUN-2003.

XX

PF 23-JUL-2001; 2001US-00911077.

XX

PR 23-JUL-2001; 2001US-00911077.

XX

PA (BLAK/) BLAKELY R D.

PA (APPA/) APPARSUNDARAM S.  
PA (FERG/) FERGUSON S.  
XX  
PI Blakely RD, Apparsundaram S, Ferguson S;  
XX  
DR WPI; 2003-810914/76.  
XX  
PT Novel isolated polynucleotide encoding human or mouse high affinity  
PT choline transporter polypeptide, useful in gene therapy to increase  
PT cholinergic function in a cell of a patient suffering from Alzheimer's  
PT disease.  
XX  
PS Disclosure; SEQ ID NO 8; 74pp; English.  
XX  
CC The present invention relates to the isolation of polynucleotide  
CC sequences encoding human and mouse high-affinity choline transporter  
CC (hCHT and mCHT respectively), and the proteins they encode. The gene  
CC encoding hCHT is located on chromosome 2q12. The polynucleotide sequence  
CC encoding hCHT, is useful for expressing hCHT recombinantly. The hCHT  
CC polynucleotide sequence when delivered to a cell, increases cholinergic  
CC function in the cell that is in a patient having Parkinson's disease,  
CC Huntington's disease, Alzheimer's disease, schizophrenia, dysautonomia or  
CC myasthenia gravis. The hCHT antibody is useful for controlling  
CC transporter CHT proteins to the brain, and for treating the above  
CC mentioned diseases. The antibody is also useful for diagnosing the above  
CC mentioned disorders and to detect the influence of cholinergic  
CC signalling. The present sequence represents *Caenorhabditis elegans* CHOI  
CC protein. Note: The sequence data for this patent was obtained in  
CC electronic format directly from the USPTO web site at seqdata.uspto.gov.  
XX  
SQ Sequence 576 AA;

Query Match 48.9%; Score 1453; DB 7; Length 576;  
Best Local Similarity 50.5%; Pred. No. 4.7e-137;  
Matches 295; Conservative 95; Mismatches 150; Indels 44; Gaps 9;

Qy 7 GLIAIIVFYLLILLVGIWAAWRTKNSGSAEER----SEAIIVGGRDIGLLVGGFTMTATW 62  
|::||: ||:||||:|||| |::| | :| :: ||:| || | |||||  
Db 6 GIVAIVFFYVLILVVGIWAGRKSKSSKELESEAGAATEEVMLAGRNIGTLVGIFTMTATW 65

Qy 63 VGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQIY 122  
||| |||||::| | || ||:|::|::| || || :|:| ||||| |  
Db 66 VGGAYINGTAEALY--NGLLGCQAPVGYAISLVMGGLLFAKKMREEGYITMLDPFQHKY 123

Qy 123 GKRMGGLLFIPALMGEMFWAAAFSALGATISVIIDVDMHISVIISALIATLYTLVGGLY 182  
|:|:|::|::|::| || || || ||||:|::|::|: || :| || || || |  
Db 124 GQRIGGLMYVPALLGETFWTAAILSALGATLSVILGIDMNASVTLACIAVFYFTFTGGYY 183

Qy 183 SVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDS-SEVYSWL 241  
:|||||:|||||: || |: | || | |::| : | |:  
Db 184 AVAYTDVVQLFCIFVGLWVCVPAAMVHDGAKDISRNA-----GDWIGEIGGFKETSLWI 237

Qy 242 DSFLLMLGGIPWQAYFQRVLSSSSATYAQVLSFLAAGCLVMAIPAILIGAIGASTDWN 301  
| |||: ||||| |||||:| || ||:| ||:| || ||| :||  
Db 238 DCMLLLVFGGIPWQVYFQRVLSKTAHGAQTLSEFVAGVGCILMAIPALIGAIARNTDWR 297

Qy 302 QTAYGLPDPKTTTEEA-----DMILPIVLQYLCVPYISFFGLGAVSAAVMSSADSSILSA 355



	:     : : : :         : :                     :	
Db	298 MTDYSPWNNGTKVESIPDKRNMVPLVFQYLTPRWVAFIGLGAVSAAVMSSADSSVLSA	357
Qy	356 SSMFARNIYQLSFRQNASDKEIVVMRITVVFVGASATAMALLTKTVYGLWYLSSDLVYI	415
	:             : :   :   :   : :       :         : :         :       :	
Db	358 ASMFAHNIWKLTIRPHASEKEVIIVMRIAICVGIMATIMALTIQSIYGLWYLCADLVYV	417
Qy	416 VIFPQLLCVLFVKGTNTYGAVAGYVSGFLRITGGEPYLYLQPLIFYPGYPPDDNGIYNQ	475
	: :             : : :       : :         :       :   :	
Db	418 ILFPQLLCVVYMPRSNTYGSLAGYAVGLVLRLLIGGEPLVSLPAFFHYPMY---TDGV--Q	472
Qy	476 KFPFKTLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDVAVV---ARHSEENMDKTILV	532
	:       :     :   : :     :     :           :	
Db	473 YFPFRTTAMLSSMATIYIVSIQSEKLFKSGRLSPEWDVMGCVVNIPIDHVPLPSDVSEFAV	532
Qy	533 KNENIKL-----DELALVKPRQSMTLSSTFTN	559
	:   : :       :   :     :	
Db	533 SSETLNMKAPNGTPAPVHPNQQPSDENTLLHPYSDQSYYSTNSN	576

Search completed: March 22, 2004, 15:32:18  
Job time : 118 secs

OM protein - protein search, using sw model

Run on: March 22, 2004, 15:30:29 ; Search time 33 Seconds  
(without alignments)  
907.366 Million cell updates/sec

Title: US-10-069-541-6  
Perfect score: 2972  
Sequence: 1 MAFHVEGLIAIIVFYLLILL.....EAFLDVDSSPEGSGTEDNLQ 580

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	2972	100.0	580	4	US-09-657-252-2	Sequence 2, Appli
2	308.5	10.4	662	1	US-07-841-651-4	Sequence 4, Appli
3	301	10.1	494	4	US-09-252-991A-24099	Sequence 24099, A
4	298.5	10.0	675	4	US-10-162-012-27	Sequence 27, Appl
5	298	10.0	521	4	US-09-540-236-2193	Sequence 2193, Ap
6	298	10.0	672	1	US-07-841-651-2	Sequence 2, Appli
7	298	10.0	672	1	US-07-841-651-3	Sequence 3, Appli
8	292	9.8	672	4	US-10-162-012-30	Sequence 30, Appl
9	277.5	9.3	548	4	US-09-543-681A-4994	Sequence 4994, Ap
10	274.5	9.2	554	4	US-09-543-681A-6886	Sequence 6886, Ap
11	272.5	9.2	718	4	US-09-657-960-3	Sequence 3, Appli

12	262.5	8.8	518	4	US-09-134-001C-4744	Sequence 4744, Ap
13	259	8.7	501	4	US-09-328-352-6371	Sequence 6371, Ap
14	257.5	8.7	551	4	US-09-252-991A-27829	Sequence 27829, A
15	255	8.6	508	4	US-09-489-039A-7541	Sequence 7541, Ap
16	247	8.3	565	4	US-09-489-039A-8414	Sequence 8414, Ap
17	243.5	8.2	552	4	US-09-252-991A-29652	Sequence 29652, A
18	241	8.1	591	4	US-09-540-236-2490	Sequence 2490, Ap
19	236.5	8.0	486	4	US-09-543-681A-6238	Sequence 6238, Ap
20	235	7.9	447	4	US-10-162-012-29	Sequence 29, Appl
21	232.5	7.8	465	4	US-09-198-452A-524	Sequence 524, App
22	231	7.8	643	4	US-09-640-198D-2	Sequence 2, Appli
23	230	7.7	496	4	US-09-489-039A-8772	Sequence 8772, Ap
24	218	7.3	506	4	US-09-543-681A-8246	Sequence 8246, Ap
25	218	7.3	618	4	US-08-595-553A-2	Sequence 2, Appli
26	218	7.3	618	4	US-09-640-198D-4	Sequence 4, Appli
27	206.5	6.9	530	4	US-09-134-001C-4510	Sequence 4510, Ap
28	206.5	6.9	644	4	US-09-252-991A-21730	Sequence 21730, A
29	205.5	6.9	497	4	US-09-543-681A-6926	Sequence 6926, Ap
30	200	6.7	635	2	US-09-014-969-11	Sequence 11, Appl
31	198	6.7	493	4	US-09-489-039A-13709	Sequence 13709, A
32	185	6.2	585	4	US-09-328-352-6133	Sequence 6133, Ap
33	138.5	4.7	467	4	US-09-107-532A-6175	Sequence 6175, Ap
34	138	4.6	482	4	US-09-134-001C-4309	Sequence 4309, Ap
35	136	4.6	345	4	US-09-489-039A-9471	Sequence 9471, Ap
36	132	4.4	499	4	US-09-252-991A-23328	Sequence 23328, A
37	129	4.3	419	4	US-09-543-681A-4862	Sequence 4862, Ap
38	128	4.3	996	4	US-09-252-991A-27018	Sequence 27018, A
39	126.5	4.3	499	4	US-09-489-039A-8830	Sequence 8830, Ap
40	126.5	4.3	670	4	US-09-134-000C-4606	Sequence 4606, Ap
41	123.5	4.2	657	4	US-09-252-991A-27682	Sequence 27682, A
42	123	4.1	496	4	US-09-134-001C-3001	Sequence 3001, Ap
43	122.5	4.1	526	4	US-09-134-000C-4715	Sequence 4715, Ap
44	121.5	4.1	538	4	US-09-107-532A-5563	Sequence 5563, Ap
45	120	4.0	485	4	US-09-198-452A-1105	Sequence 1105, Ap

#### ALIGNMENTS

##### RESULT 1

US-09-657-252-2

; Sequence 2, Application US/09657252

; Patent No. 6500643

; GENERAL INFORMATION:

; APPLICANT: Wu, Dong-Hai

; APPLICANT: Gu, Yunrong

; APPLICANT: Millard, William

; APPLICANT: He, Yun-Je

; TITLE OF INVENTION: Human High Affinity Choline Transporter cDNA

; FILE REFERENCE: MBHB00-639

; CURRENT APPLICATION NUMBER: US/09/657,252

; CURRENT FILING DATE: 2000-09-07

; NUMBER OF SEQ ID NOS: 6

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 2

; LENGTH: 580

; TYPE: PRF

; ORGANISM: Homo sapiens  
US-09-657-252-2

Query Match 100.0%; Score 2972; DB 4; Length 580;  
Best Local Similarity 100.0%; Pred. No. 1.6e-281;  
Matches 580; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTA 60
        |||
Db      1 MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTA 60

Qy     61 TWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ 120
        |||
Db     61 TWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ 120

Qy    121 IYGKRMGGLLFIPALMGEMFWAAAFSALGATISVIIDVDMHISVIISALIATLYTLVGG 180
        |||
Db    121 IYGKRMGGLLFIPALMGEMFWAAAFSALGATISVIIDVDMHISVIISALIATLYTLVGG 180

Qy    181 LYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSSEVYSW 240
        |||
Db    181 LYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSSEVYSW 240

Qy    241 LDSFLLMLGGIPWQAYFQRVLSSSSATYAQVLSFLAAFGLVMAIPAILIGAIGASTDW 300
        |||
Db    241 LDSFLLMLGGIPWQAYFQRVLSSSSATYAQVLSFLAAFGLVMAIPAILIGAIGASTDW 300

Qy    301 NQTAYGLPDPKTTEEADMILPIVLQYLCVPYISFFGLGAVSAAVMSSADSSILSASSMFA 360
        |||
Db    301 NQTAYGLPDPKTTEEADMILPIVLQYLCVPYISFFGLGAVSAAVMSSADSSILSASSMFA 360

Qy    361 RNIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQ 420
        |||
Db    361 RNIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQ 420

Qy    421 LLCVLFVKGTNTYGAVAGYVSGFLRITGGEPYLYLQPLIFYPGYYPDDNGIYNQKFPFK 480
        |||
Db    421 LLCVLFVKGTNTYGAVAGYVSGFLRITGGEPYLYLQPLIFYPGYYPDDNGIYNQKFPFK 480

Qy    481 TLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDVAVVARHSEENMDKTILVKKNENIKLD 540
        |||
Db    481 TLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDVAVVARHSEENMDKTILVKKNENIKLD 540

Qy    541 ELALVKPRQSMTLSSTFTNKEAFLDVDSSPEGSGTEDNLQ 580
        |||
Db    541 ELALVKPRQSMTLSSTFTNKEAFLDVDSSPEGSGTEDNLQ 580
```

RESULT 2

US-07-841-651-4

; Sequence 4, Application US/07841651

; Patent No. 5410031

; GENERAL INFORMATION:

; APPLICANT: Pajor, Ana M

; APPLICANT: Wright, Ernest M

; TITLE OF INVENTION: Cloning and Functional Expression of a

```

; TITLE OF INVENTION: Mammalian Na+/Nucleoside Cotransporter: A Member of
the
; TITLE OF INVENTION: SGLT Family
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheldon & Mak
; STREET: 225 South Lake Avenue, Ninth Floor
; CITY: Pasadena
; STATE: California
; COUNTRY: USA
; ZIP: 91101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/841,651
; FILING DATE: 19920224
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandel, SaraLynn
; REGISTRATION NUMBER: 31,853
; REFERENCE/DOCKET NUMBER: 8772
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (818) 796-4000
; TELEFAX: (818) 795-6321
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 662 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Oryctolagus cuniculus
US-07-841-651-4

```

```

Query Match          10.4%; Score 308.5; DB 1; Length 662;
Best Local Similarity 23.4%; Pred. No. 3.2e-21;
Matches 154; Conservative 110; Mismatches 238; Indels 155; Gaps 26;

```

```

Qy      11 IIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGYING 70
      |:::|::: ||:| | : | | | :      : | | :      | :: | : | | : |
Db      32 IVIYFLVVMVAVGLWAMFST-NRGTV----GGFFLAGRSMVWWPIGASLFASNIGSGHFVG 86

Qy      71 TAEAVYVPGYGLAWAQAPIGYS-----LSLILGGLFFAKPMRSKGYVTMLDPFQQIY-GK 124
      |      | | | | | | :      : ::| | :|      :| : | || | : | : : | |
Db      87 IA-----GTGAASGIATGGFEWNALIMVVVLGWVFPVIYIRA-GVVTMPEYLQKRFGGK 139

Qy      125 RMGGLLFIPALMGEMFW--AAAFSALGAT-ISVIIDVDMHISVIISALIATLYTLVGGL 181
      | :      | | :| : | : | | | | | : : :|:::| | :| | | : | | |
Db      140 RIQIYLSILSLLLYIFTKISADIFS--GAIFIQLTLGLDIYVAIILLVITGLYTITGGL 197

Qy      182 YSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKY-----Q 225
      :| | | :| : | | | | | | : | | | |
Db      198 AAVIYDTLQTAIMMVGSVILTGFAFHEVG----GYEAFTEKYMRAIPSQISYGNTSIPQ 253

```

Qy 226 KPWLGTVDSSSEVYSWLDLSFLLMLGGIPW-----QAYFQRVLSSSSA 267  
| : | : : | : | || | | : :  
Db 254 KCYTPREDAFHI-----FRDAITGDIPWPGLVFGMSILTLWYWCTDQVIVQRCLSAKNL 307  
Qy 268 TYAQVLSFLAAGCLVMAIPAIIIGAIGASTDWNQTAYGLPDP-----KTTEEADMILP 321  
: : : | : : : : : | : : | : : : :  
Db 308 SHVKAGCILCGYLKVMMPFLIVMMGMVSRILYTDKVACVVPSECERYCGTRVGCTNIAFP 367  
Qy 322 IVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFARNIYQLSFRQNASDKEIVWVM 381  
: : | | : | : | : | | | | : | | : : : :  
Db 368 TLVVELMPNGLRGLMLSVMMASLMSLTSIFNSASTLFTMDIY-TKIRKKASEKELMIAG 426  
Qy 382 RI-TVFVFGASATAMALLTKTVYG--LWYLSSDLVYI--VIFPQLLCVLFVKGNTNTYGAV 436  
| : : | : | : : : | : | : | : | : | : | :  
Db 427 RLFMLFLIGISIAWVPIVQSAQSGQLFDYIQSITSYLGPPIAAVFLIAIFWKRVNEPGAF 486  
Qy 437 AGYVSGFLFLRI-----TG-----GEPYLYLQPLIFYPGYYPDDNGIY 473  
| | | : | | | | | : :  
Db 487 WGLVLGFLIGISRMITEFAYGTGSCMEPSNCPTIICGVHYLYFAIILF----- 534  
Qy 474 NQKFPFKTLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDVAVVA-RHSEENMDKTILV 532  
| | : | : : | | : | : : : | : : :  
Db 535 -----VISIITVVVVSFLTCKPI-----PDVHLYRLCWSLRNSKE----- 568  
Qy 533 KNENIKLD--ELALVKPRQSMTLSSTFTNKEAF-----LDVDSSPEGSGTED 577  
| | || | : : : | : | : | | : : :  
Db 569 --ERIDLAGEEDIQEAPEEATDTEVPKKGFFRAYDLFCGLDQDKGPKMTKEEE 623

# RESULT 3

US-09-252-991A-24099

; Sequence 24099, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 24099

; LENGTH: 494

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

; FEATURE:

; NAME/KEY: UNSURE

; LOCATION: (232)

; OTHER INFORMATION: Identity of amino acid at the above locations are unknown.

US-09-252-991A-24099

Query Match 10.1%; Score 301; DB 4; Length 494;  
 Best Local Similarity 24.9%; Pred. No. 1.1e-20;  
 Matches 116; Conservative 82; Mismatches 207; Indels 60; Gaps 15;

```

Qy      9 IAIIVFYLLILLVGI----WAAWRTKNSGSAERSEAIIVGGRDIGLLVGGF---TMTAT 61
      |: :| :|| |: | | |: | : | ||::| || || ||
Db     32 MALDIFVVLIIYAAGMIALGWYGMR-----RAKTRDD-YLVAGRNLG---PGFYLGTMAT 82

Qy     62 WVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQI 121
      :|| || || | | || | :| ||||: |: ::
Db     83 VLGGASTIGTVRLGYVHGISGFWLCAIG--LGIVGLSLFLAKPLLKLIYTVTQVLER 140

Qy    122 YGKRMGGLLFIPALMGEMFWAAAI FSALGATISVIIDVDMHISVIISALIATLYTLVGGL 181
      | : |: : | |:| : |: : |:|:| : ||: ||:
Db    141 YNPAARHASALIMLVYALMIGATSTIAIGTVMQVLFGLPFWVSILIGGGVVVLYSTIGGM 200

Qy    182 YSVAYTDVVQLFCIFVGL-WISVPFALSHPAVAD-----IGFTAVHAKYQKPWL 230
      |: ||:| |: ||| :: :| ::| | ||: |
Db    201 WSLTLTDIVQFLIMTVGLVFLMPLSINDAGXWDALVAKLPASYFDFTAI-----GW-- 252

Qy    231 TVDSSEVYSWLDSFLLLMLGGIPWQAYFQRVLSSSSATYAQVLSFLAAGCLVMAIPAIL 290
      |: | ||: | | :||| :: | | |:| | |:: : |
Db    253 --DTIVTY-----FLIYFFGIFIGQDIWQRVFTARSETVAKVAGSAAGIYCVLYGMAGAL 305

Qy    291 IGAIGASTDWNQTAYGLPDKTTEEADMILPIVLQYLCVPYISFFGLGAVSAAVMSSADS 350
      || || | : |:: | | : | ||:|:| :
Db    306 IGMAAKVL-----LPD---LENVNNAFASVVEHSLPNGIRGLVIAAALAALMSTASA 354

Qy    351 SILSASSMFARNIY-QLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKTVYGLWYLS 409
      :|:|: :: :| : || | || :| :||: | ::
Db    355 GLLAASTTVTQDLLPRLRRGRGQSDNGDVHENRIATLLLGLVVLGIALVVSDDVISALTVA 414

Qy    410 SDLVYIVIFPQLLCVLFVKGTNTYGAVA----GYVSGFLRLRITGG 450
      |: : |: :: | | ||: |:: | | |
Db    415 YNLLVGMLIPLIGAIYWKRATTAGAITSMTLGFLTVLVFMIKDG 459

```

#### RESULT 4

US-10-162-012-27

; Sequence 27, Application US/10162012

; Patent No. 6682597

; GENERAL INFORMATION:

; APPLICANT: Curtis, Rory A.J.

; APPLICANT: Silos-Santiago, Inmaculada

; APPLICANT: Gu, Wei

; TITLE OF INVENTION: NOVEL HUMAN ION CHANNEL AND TRANSPORTER FAMILY MEMBERS

; FILE REFERENCE: 10448-190001

; CURRENT APPLICATION NUMBER: US/10/162,012

; CURRENT FILING DATE: 2002-06-04

; PRIOR APPLICATION NUMBER: US 60/209,845

; PRIOR FILING DATE: 2000-06-06

; PRIOR APPLICATION NUMBER: US 09/875,321

; PRIOR FILING DATE: 2001-06-06

; PRIOR APPLICATION NUMBER: PCT/US01/18340

; PRIOR FILING DATE: 2001-06-06

; PRIOR APPLICATION NUMBER: US 60/209,257

```

; PRIOR FILING DATE: 2000-06-05
; PRIOR APPLICATION NUMBER: US 09/875,423
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: PCT/US01/18398
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/209,238
; PRIOR FILING DATE: 2000-06-05
; PRIOR APPLICATION NUMBER: US 09/875,363
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: PCT/US01/18247
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/227,068
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: US 09/928,530
; PRIOR FILING DATE: 2001-08-13
; PRIOR APPLICATION NUMBER: PCT/US01/25475
; PRIOR FILING DATE: 2001-08-15
; PRIOR APPLICATION NUMBER: US 60/226,770
; PRIOR FILING DATE: 2000-08-21
; PRIOR APPLICATION NUMBER: US 09/934,421
; PRIOR FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: PCT/US01/26096
; PRIOR FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: US 60/279,281
; PRIOR FILING DATE: 2001-03-28
; PRIOR APPLICATION NUMBER: US 10/109,029
; PRIOR FILING DATE: 2002-03-28
; PRIOR APPLICATION NUMBER: PCT/US02/09728
; PRIOR FILING DATE: 2002-03-28
; PRIOR APPLICATION NUMBER: US 60/290,288
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US (not assigned)
; PRIOR FILING DATE: 2002-05-13
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 27
;   LENGTH: 675
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-162-012-27

```

```

Query Match          10.0%;  Score 298.5;  DB 4;  Length 675;
Best Local Similarity 22.7%;  Pred. No. 3.2e-20;
Matches 149;  Conservative 115;  Mismatches 238;  Indels 155;  Gaps 29;

```

```

Qy      2 AFHVEGL----IAIIVFY-LLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGF 56
      || :||  ||::| | | :| ||::| : ||      :      : | :  ||
Db      18 AFPQKGLEPGDIAVLVLYFLFVLAVGLWSTVKTGR---DTVKGYFLAEGNMVWWPVGA- 72

Qy      57 TMTATWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLS-----LILGGLFFAKPMRSKGY 111
      :: | : || | : | |      | | | : | | : | : | : | | : |
Db      73 SLFASNVGSGHFGLA-----GSGAATGISVSAYELNGLFSVLMWLAWIFL--PIYIAGQ 124

Qy      112 VTMLDPFQQIYGKRMGGLLFIPALMGEMFWAAAI FSALGATI---SVID---VDMHIS 164
      || : : :  || ||: || :: : :  ||: : : : : | : : : :
Db      125 VTTMPEYLR--KRFGGIR-IPILAVLYLFYIYFTKISVDMYAGAI FIQSSHLDLYLA 180

```



Qy 165 VIISALIATLYTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKY 224  
 Db 181 IVGLLAITAVYTVAGGLAAVIYTDALQTLIMLIGALTLMGY--SFAAVG--GMEGLKEKY 236

Qy 225 -----QKPWLGTVDSSSEVYS-WLDSFLLMLGGI 252  
 Db 237 FLALASNRSENSSCGLPREDAFHIFRDPLTSDLPPWPGVLFGMSIPSLWY----- 285

Qy 253 PW---QAYFQRVLSSSSATYAQVLSFLAAGCLVMAIPAILIGAIGASTDWNQTAYGLPD 309  
 Db 286 -WCTDQVIVQRTLAAKNLSHAKGALMAAYLKVLPFIMVFPGMVSRILFPDQVA--CAD 342

Qy 310 PKTTEE-----ADMILPIVLQYLCVPYISFFGLGAVSAAMVSSADSSILSASSMFAR 361  
 Db 343 PEICQKICSNPSGCSDIAYPKLVLELLPTGLRGLMMAMVMAALMSSSLTSIFNSASTIFTM 402

Qy 362 NIYQLSFRQNASDKEIVWVMRITVVFVGASATAMALLTKTVYGLW-----Y 407  
 Db 403 DLWN-HLRPRASEKELMIVGRVFV-----LLLVLVSILWIPVVQASQGGQLFIY 450

Qy 408 LSSDLVYI----VIFPQLLCVLFVKGNTNTYGAVAGYVSGFLFRITG-GEPLYLQPLIF 461  
 Db 451 IQSISSYLQPPVAVVF---IMGCFWKRTNEKGAFWGLISGLLLGLVRLVLDIFIYVQPRC- 506

Qy 462 YPGYYPDDNGIYNQKFPFKTLAMVTSFLTNICISYLAKYLFESGTLPPKLDV----- 513  
 Db 507 ---DQDERPVLVKSIIHYLYFSMILSTVTLITVSTVSWF-----TEPPSKEMVSHLTWFT 558

Qy 514 -FDAVVARHSEENMDKTIIVKNENIKLD-----ELALVKPRQSMSTLSSTFTNKEA 562  
 Db 559 RHDVQVQEQAPPAAPLSLTLSQNGMPEASSSSSVQFEMVQENTSKTHSCDMTPKQS 615

RESULT 5

US-09-540-236-2193

; Sequence 2193, Application US/09540236

; Patent No. 6673910

; GENERAL INFORMATION:

; APPLICANT: Gary L. Breton et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARRHALIS

; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 2709.2005-001

; CURRENT APPLICATION NUMBER: US/09/540,236

; CURRENT FILING DATE: 2000-04-04

; NUMBER OF SEQ ID NOS: 3840

; SEQ ID NO 2193

; LENGTH: 521

; TYPE: PRT

; ORGANISM: M.catarrhalis

US-09-540-236-2193

Query Match 10.0%; Score 298; DB 4; Length 521;

Best Local Similarity 23.4%; Pred. No. 2.4e-20;

Matches 131; Conservative 103; Mismatches 212; Indels 114; Gaps 20;

Qy 9 IAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGYI 68



```

; COUNTRY: USA
; ZIP: 91101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/841,651
; FILING DATE: 19920224
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandel, SaraLynn
; REGISTRATION NUMBER: 31,853
; REFERENCE/DOCKET NUMBER: 8772
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (818) 796-4000
; TELEFAX: (818) 795-6321
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 672 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-841-651-2

```

```

Query Match          10.0%; Score 298; DB 1; Length 672;
Best Local Similarity 25.0%; Pred. No. 3.5e-20;
Matches 153; Conservative 89; Mismatches 232; Indels 138; Gaps 25;

```

```

Qy      9 IAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGY 67
        ||:| :||:: ||::| || | | :          : || :   | :: | : | | :
Db      26 IAVIAAYFLLVIGVGLWSMCRT-NRGTV----GGYFLAGRSMVWVPVGASLFASNIGSGH 80

Qy      68 INGTAEAVYVPGYGLAWAQAPIGYSLS----LILGGLFFAKPMRSKGYVTMLDPFQQIYG 123
        | |      | | | | | | : :      : : | | | | : | : ||      |
Db      81 FVGLA-----GTGAANGLAVAGFEWNALFVLLLLGWLFAPVYLTAGVITM----PQYLR 130

Qy      124 KRMGG-----LLFIPALMGEMFWAAAI--SALGATISVIIDVDMHISVIISA 169
        || ||      | : | :      : : | : | | | | : | | |
Db      131 KRFGGHRIRLYLSVLSLFLYIFTKISVDMFSGAVFIQQALGWNI-----YASVIAL 182

Qy      170 LIATLYTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKY----- 224
        | : || : || | : || | | | | | : | : | : | : | : | : |
Db      183 GITMVYTVTGGLAALMYTDTVQTFVIIAGAFILTGYAFHEVG----GYSGLFDKYMGA 238

Qy      225 -----QKPWLGTVDSSSEVYSWLDSFLLL---MLGGIPW-----QAYF 258
        : | : | : ||      || : || : | : ||      |
Db      239 SLTVSEDPVGNISSSCYRPRPDSYHLLRDPVTGDLPPALLLGLTIVSGWYWCSDQVIV 298

Qy      259 QRVLSSSSATYAQVLSFLAAGCLVMAIPAILIGAIGASTDWNQTAYGLPDPKT-----TE 314
        || | : : | : : | : | : : | | : : | | : | : | |
Db      299 QRCLAGRNLTTHIKAGCILCGYLKLTMPFLMVMPGMISRILYPDEVACVAPEVCKRVC 358

Qy      315 E--ADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFARNIYQLSFRQNA 372
        : : : | : : | | : | : | : || | | | | : | : | | |
Db      359 VGCSNIAYPRLVVKLMPNGLRGLMLAVMLAALMSSLASIFNSSSTLFTMDIYTL--RPRA 416

```

Qy 373 SDKEIVVWMRITVFVFGASATAMALLTKTVYG---LWYLSSDLVYIV--IFPQLLCVLFV 427  
: |:: | | : | : | : : | : | | : : : |||  
Db 417 GEGELLLVGRWLWVVFIVAVSVAWLPVVQAAQGGQLFDYIQSVSSYLAPPVSAVFVVALFV 476  
Qy 428 KGTNTYGAVAGYVSGFLFRITGGEPYLYLQPLIFYPGYPPDDNGIYNQKFPFKTLAMV-- 485  
| || | : || : : | : | | : |  
Db 477 PRVNEKGAFWGLIGLLMGLARLIP-----EFSFGTGSCVRP 513  
Qy 486 ---TSFLTNICISYLAKYLFE-SG-----TLP-PKLDVFDVAVVA-RHSEENMDKTI 530  
: || : | | || || | | | : : | : ||| : |  
Db 514 SACPAFLCRVHYLYFAIVLFFCSGLLIIIVSLCTAPIPRKHLHRLVFSRLHSKE----- 567  
Qy 531 LVKNENIKLDEL 542  
: |:: |||  
Db 568 --EREDLDADEL 577

RESULT 7

US-07-841-651-3

; Sequence 3, Application US/07841651

; Patent No. 5410031

; GENERAL INFORMATION:

; APPLICANT: Pajor, Ana M

; APPLICANT: Wright, Ernest M

; TITLE OF INVENTION: Cloning and Functional Expression of a

; TITLE OF INVENTION: Mammalian Na+/Nucleoside Cotransporter: A Member of  
the

; TITLE OF INVENTION: SGLT Family

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sheldon & Mak

; STREET: 225 South Lake Avenue, Ninth Floor

; CITY: Pasadena

; STATE: California

; COUNTRY: USA

; ZIP: 91101

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/841,651

; FILING DATE: 19920224

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Mandel, SaraLynn

; REGISTRATION NUMBER: 31,853

; REFERENCE/DOCKET NUMBER: 8772

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (818) 796-4000

; TELEFAX: (818) 795-6321

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 672 amino acids

; TYPE: AMINO ACID

```
;      TOPOLOGY: linear
;      MOLECULE TYPE: protein
;      HYPOTHETICAL: NO
;      ORIGINAL SOURCE:
;      ORGANISM: Oryctolagus cuniculus
US-07-841-651-3
```

```
Query Match          10.0%; Score 298; DB 1; Length 672;
Best Local Similarity 25.0%; Pred. No. 3.5e-20;
Matches 153; Conservative 89; Mismatches 232; Indels 138; Gaps 25;
```

```
Qy      9 IAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGY 67
      ||:| ::||:: ||:|: || | |:      : || : | :: |: :| |:
Db      26 IAVIAAYFLLVIGVGLWSMCRN-NRGTV----GGYFLAGRSMVWVPVGASLFASNIGSGH 80

Qy      68 INGTAEEAVYVPGYGLAWAQAPIGYSLS----LILGGLFFAKPMRSKGYVTMLDPFQQIYG 123
      | |      | | | | |: : ::| | || : | :|| |
Db      81 FVGLA-----GTGAANGLAVAGFEWNALFVLLLLGWLFAPVYLTAGVITM----PQYLR 130

Qy      124 KRMGG-----LLFIPALMGEMFWAAAI--SALGATISVIIDVDMHISVIISA 169
      || ||      |:| : :: |:| || | | : |||
Db      131 KRFGGHRIRLYLSVLSLFLYIFTKISVDMFSGAVFIQQALGWN-----YASVIAL 182

Qy      170 LIATLYTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKY----- 224
      | :||: ||| :: ||| || | | |:| :| :| : : ||
Db      183 GITMVTVTGGLAALMYTDTVQTFVIIAGAFILTGYAFHEVG----GYSGLFDKYMAMT 238

Qy      225 -----QKPWLGTVDSSSEVYSWLDSFLLL---MLGGIPW-----QAYF 258
      : | :| : || ||: || : | :|| |
Db      239 SLTVSEDPVGNISSSCYRPRPDSYHLLRDPVTGDLWPALLLGLTIVSGWYWCSDQVIV 298

Qy      259 QRVLSSSSATYAQVLSFLAAFGCLVMAIPAILIGAIGASTDWNQATAYGLPDPKT----TE 314
      || |: : |: : | : | :: | | :: | |: ||
Db      299 QRCLAGRNLTTHIKAGCILCGYLKLTMPFLMVMPGMISRIYPDEVACVAPEVCKRVCSTE 358

Qy      315 E--ADMILPIVLQYLCVPYISFFGLGAVSAAVMSSADSSILSASSMFARNIYQLSFRQNA 372
      :: | :: | | : | : ||: || | |:|:| :|| | | |
Db      359 VGCSNIAYPRLVVKLMPNGLRGLMLAVMLAALMSSLASIFNSSSTLFTMDIYTL--RPRA 416

Qy      373 SDKEIVWVMRITVVFVGASATAMALLTKTVYG---LWYLSSDLVYIV--IFPQLLCVLFV 427
      : |:| | |: | | : | : : | |: | |: : : |||
Db      417 GEGELLVGRWVVFIVAVSVAWLPVVQAAQGGQLFDYIQSVSSYLAPPVSAVFVVALFV 476

Qy      428 KGTNTYGAAGYVSGFLRITGGEPYLYLQPLIFYPGYPPDDNGIYNQKFPFKTLAMV-- 485
      | || | : || : : | :| | | : |
Db      477 PRVNEKGAFWGLIGGLLMGLARLIP-----EFSEGTGSCVRP 513

Qy      486 ---TSFLTNICISYLAKYLFE-SG-----TLP-PKLDVFDVAVVA-RHSEENMDKTI 530
      :|| : | | || || | | |: : | : |||:|
Db      514 SACPAFLCRVHYLYFAIVLFFCSGLLIIIVSLCTAPIPRKHLHRLVFSLRHSKE----- 567

Qy      531 LVKNENIKLDEL 542
      : |: : |||
Db      568 --EREDLDADEL 577
```

RESULT 8

US-10-162-012-30  
; Sequence 30, Application US/10162012  
; Patent No. 6682597  
; GENERAL INFORMATION:  
; APPLICANT: Curtis, Rory A.J.  
; APPLICANT: Silos-Santiago, Inmaculada  
; APPLICANT: Gu, Wei  
; TITLE OF INVENTION: NOVEL HUMAN ION CHANNEL AND TRANSPORTER FAMILY MEMBERS  
; FILE REFERENCE: 10448-190001  
; CURRENT APPLICATION NUMBER: US/10/162,012  
; CURRENT FILING DATE: 2002-06-04  
; PRIOR APPLICATION NUMBER: US 60/209,845  
; PRIOR FILING DATE: 2000-06-06  
; PRIOR APPLICATION NUMBER: US 09/875,321  
; PRIOR FILING DATE: 2001-06-06  
; PRIOR APPLICATION NUMBER: PCT/US01/18340  
; PRIOR FILING DATE: 2001-06-06  
; PRIOR APPLICATION NUMBER: US 60/209,257  
; PRIOR FILING DATE: 2000-06-05  
; PRIOR APPLICATION NUMBER: US 09/875,423  
; PRIOR FILING DATE: 2001-06-05  
; PRIOR APPLICATION NUMBER: PCT/US01/18398  
; PRIOR FILING DATE: 2001-06-05  
; PRIOR APPLICATION NUMBER: US 60/209,238  
; PRIOR FILING DATE: 2000-06-05  
; PRIOR APPLICATION NUMBER: US 09/875,363  
; PRIOR FILING DATE: 2001-06-05  
; PRIOR APPLICATION NUMBER: PCT/US01/18247  
; PRIOR FILING DATE: 2001-06-05  
; PRIOR APPLICATION NUMBER: US 60/227,068  
; PRIOR FILING DATE: 2000-08-22  
; PRIOR APPLICATION NUMBER: US 09/928,530  
; PRIOR FILING DATE: 2001-08-13  
; PRIOR APPLICATION NUMBER: PCT/US01/25475  
; PRIOR FILING DATE: 2001-08-15  
; PRIOR APPLICATION NUMBER: US 60/226,770  
; PRIOR FILING DATE: 2000-08-21  
; PRIOR APPLICATION NUMBER: US 09/934,421  
; PRIOR FILING DATE: 2001-08-21  
; PRIOR APPLICATION NUMBER: PCT/US01/26096  
; PRIOR FILING DATE: 2001-08-21  
; PRIOR APPLICATION NUMBER: US 60/279,281  
; PRIOR FILING DATE: 2001-03-28  
; PRIOR APPLICATION NUMBER: US 10/109,029  
; PRIOR FILING DATE: 2002-03-28  
; PRIOR APPLICATION NUMBER: PCT/US02/09728  
; PRIOR FILING DATE: 2002-03-28  
; PRIOR APPLICATION NUMBER: US 60/290,288  
; PRIOR FILING DATE: 2001-05-11  
; PRIOR APPLICATION NUMBER: US (not assigned)  
; PRIOR FILING DATE: 2002-05-13  
; NUMBER OF SEQ ID NOS: 48  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 30  
; LENGTH: 672  
; TYPE: PRT  
; ORGANISM: Homo sapiens

US-10-162-012-30

Query Match 9.8%; Score 292; DB 4; Length 672;  
Best Local Similarity 24.1%; Pred. No. 1.4e-19;  
Matches 147; Conservative 91; Mismatches 237; Indels 136; Gaps 22;

```
Qy      8 LIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGY 67
      :: | ::||: ||:|: || | |:      : || : | :: |: :| |:
Db      26 ILVIAAYFLLVIGVGLWSMCRT-NRGTV---GGYFLAGRSMVWWPVGASLFASNIGSGH 80

Qy      68 INGTAEEAVYVPGYGLAWAQAPIGYSLs----LILGGLFFAKPMRSKGYVTMLDPFQQIYG 123
      | |      | | | | |: : ::| | || : | :|| |
Db      81 FVGLA-----GTGAASGLAVAGFEWNALFVLLLGWLFAPVYLTAGVITM----PQYLR 130

Qy      124 KRMGG-----LLFIPALMGEMFWAAAI--SALGATISVIIDVDMHISVIISA 169
      || ||      |:| : :: |:| ||| | : |||
Db      131 KRFGGRRIRLYLSVLSFLYIFTKISVDMFSGAVFIQQALGWN-----YASVIAL 182

Qy      170 LIATLYTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKY----- 224
      | :||: ||| :: ||| || | | | : :| :| : : ||
Db      183 GITMIYTVTGGLAALMYTDTVTQTFVILGGACILMGYAFHEVG----GYSGLFDKYLGAAT 238

Qy      225 -----QKPWLGTVDSSSEVYSWLDsFLLL---MLGGIPW-----QAYF 258
      : | :| : | ||: || : | :|| |
Db      239 SLTVSEDPVGNISSFCYRPRPDSYHLLRHPVTGDLWPALLLGLTIVSGWYWCSDQVIV 298

Qy      259 QRVLSSSSATYAQVLSFLAAGCLVMAIPAILIGAIGASTDWNQTAYGLPDPKT----TE 314
      || |: | |: : | : | :: | | :: | :| : ||
Db      299 QRCLAGKSLTHIKAGCILCGYLKLTMPFLMVMPGMISRILYPDEVACVVPEVCRRVCGETE 358

Qy      315 E---ADMILPIVLQYLCVPYISFFGLGAVSAVMSSADSSILSASSMFARNIYQLSFRQNA 372
      ::: | :: | | : | : ||: || | | :|::| :|| | |
Db      359 VGCSNIAYPRLVVKLMPNGLRGLMLAVMLAALMSSLASIFNSSSTLFTMDIY-TRLRPRA 417

Qy      373 SDKEIVWVMRI-TVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQLLCV----LFV 427
      |:|:: | |: ||: | : :: | : | : | : | |||
Db      418 GDRELLLVGRLWVVFIVVVSVAWLPVVQAAQGGQLFDYIQAVSSYLAPPVSAVFVIALFV 477

Qy      428 KGTNTYGAAGYVSGFLRLITGGEPYLYLQPLIFYPGYYPDDNGIYNQKFPFKTLAMV-- 485
      | || | : || : : | :| | : : |
Db      478 PRVNEQGAFWGLIGGLMGLARLIP-----EFSFGSGSCVQP 514

Qy      486 ---TSFLTNICISYLAKYLFE-SGTLPPKLDVFDVAV-----ARHSEENMDKTI 530
      :|| : | | || || | : : | : ||:|
Db      515 SACPAFLCGVHYLYFAIVLFFCSGLLTLTVSLCTAPIPRKHLHRLVFSRLHSKE----- 568

Qy      531 LVKNENIKLDE 541
      : |:: ||
Db      569 --EREDLDADE 577
```

RESULT 9

US-09-543-681A-4994

; Sequence 4994, Application US/09543681A

; Patent No. 6605709

; GENERAL INFORMATION:

; APPLICANT: GARY BRETON

```
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS
MIRABILIS FOR
; TITLE OF INVENTION:  DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 4994
; LENGTH: 548
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-4994
```

```
Query Match          9.3%; Score 277.5; DB 4; Length 548;
Best Local Similarity 24.3%; Pred. No. 2.6e-18;
Matches 144; Conservative 95; Mismatches 229; Indels 125; Gaps 23;
```

```
Qy      4 HVEGL-----IAIIVFYLLILL-VGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGF 56
      | ||:      ||  |::|:::|::|      ::  |:::  :  |::  |
Db      6 HTEGVGLSTIDYAI FALYVIIIISLGLWV---SRSKDGAKKGTKDYFLAGKTL PWWAIGS 62

Qy     57 TMTATWV-----GGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMR 107
      :: | :      | | : | | | |      | |      |::|:  :|
Db     63 SLIAANISAEQFIGMSGSGFSIGLAIASY-----EWMAA-----LTLIIVAKYFLPIFI 111

Qy    108 SKGYVTMLDPFQQIYGKRMGGLLFIPALMGEMFWAAA-IFSAL-----GATISVIIDV 159
      || |:: : : : | | | | :|| | || |      : |:: |
Db    112 EKGIIYTIPEFVENRFKSR--NLKTLA----VFWLALFIFVNLTSVLYLGSLALETILGV 165

Qy    160 DMHISVIIISALIATLYTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSH----- 209
      | :::| || | :::| || | :||:| || |:| : : | : |::
Db    166 PMMYAIIIGLALFAVIYSLYGGLSAVAWTDVVQVFFLILGGLFTTVLAVSYIGDGGIMEG 225

Qy    210 ----PAVADIGFTAVHAKYQKPWLGTVDSSEVYSWLD SFLLLMLGGIPW-----Q 255
      | | | : ||  ::      : :::||: |      |
Db    226 LSKMTAAAPDHFKMILAKENPQFMNLPG-----IAVLIGGL-WVANLYYWGFNQ 273

Qy    256 AYFQRLVSSSSATYAQVLSFLAAGCLVMAIPAIL--IGAIGASTDWNQTAYGL-----P 308
      || |:: | ||  || | ::: | :: | | :|| | ||
Db    274 YIIQRALAAKSINEAQKGLVFAAFLKLIVPILVVPGIAAFVITTDPTLMA-GLGTMAQE 332

Qy    309 DPKTTEEADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFARNIYQLSF 368
      | :|| | : |::| | : :||:| || | : | :::| :||:
Db    333 HIPTLAQADKAYPWL TQFL-PIGAKGVVFAALAAAVSSLASMLNSIATIFTMDIYKEYI 391

Qy    369 RQNASDKEIVWVMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYI-----VIFPQLLC 423
      :|: :| | ||: : | :| |      | : ||      : : | :|
Db    392 GPKSSETRLVNVGRISAVIALIIACFIAPL-----LGGIDQAFQYIQEYTG LVSPGILA 445

Qy    424 V----LFVKGTNTYGAVAGYVSGLEFLRITGGEPYLYLQPLIFYPGYPPDDNGIYNQKFPF 479
      | || | || ||: | | : :      : | | ||      ||
Db    446 VFLLGLEFWKKTNAKGAIIGVVLSIPFAL-----FLKLMPL-----GMPF 484

Qy    480 KTLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDVAVVARHSEENMDKTILV 532
      | | | : | ::  : |      : : ||      | | : |::
```



Db 485 LDQMMYTFIFTAVVIGLVSLTSTKSDSVGAIVLTDATEFKTQSGFNIA SYIIM 537

RESULT 10

US-09-543-681A-6886

; Sequence 6886, Application US/09543681A

; Patent No. 6605709

; GENERAL INFORMATION:

; APPLICANT: GARY BRETON

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS FOR

; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 2709.1002-001

; CURRENT APPLICATION NUMBER: US/09/543,681A

; CURRENT FILING DATE: 2000-04-05

; PRIOR APPLICATION NUMBER: US 60/128,706

; PRIOR FILING DATE: 1999-04-09

; NUMBER OF SEQ ID NOS: 8344

; SEQ ID NO 6886

; LENGTH: 554

; TYPE: PRT

; ORGANISM: Proteus mirabilis

US-09-543-681A-6886

Query Match 9.2%; Score 274.5; DB 4; Length 554;

Best Local Similarity 23.0%; Pred. No. 5.1e-18;

Matches 125; Conservative 96; Mismatches 214; Indels 109; Gaps 24;

```
Qy      6 EGLIAIIVFYLLILLVGIWAAWRTKNSGSAERSEAIIVGGRDIGL-----LVGGFTMTA 60
      : :| :| | | : ||: ||: ||: ||: | : | | |
Db     38 QAIIMFLIFVGLTLYITYWASKRTRS-----RSDYYTAGGKITGFQNGMAIAGDFMSAA 91

Qy     61 TWVGGGYINGTAEAVYVPGY-GLAWAQAPIGYSLSLILGG---LFFAKPMRSKGYVTML 115
      ::: | : || || || ||: ||: ||: | : ||: | |
Db     92 SFL-----GISALVYTSGYDGLI-----YSIGFLIGWPIILFIIAERLRNLGRYTF 138

Qy    116 DPFQ-QIYGKRMGGLLFIPALMGEMFWAAAI FSALGATISVIIDVDMHISVIISALIATL 174
      | : : | : | | :| : : | | : : : ||: ||: : : |
Db    139 DVVSYRLSPKPIRTLSAIGSLVVVALYLIAQMVGAGKLIELFLGNYHIAVILVGLMVL 198

Qy    175 YTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAK----YQKPWL- 229
      | | ||: : : : : : | : : || | : : : ||:
Db    199 YVLFGGMLATTWVQIIKAILLLAGATFMAVMVMK---AADFNENTLFKEAVNVHQKGF 255

Qy    230 ----GTVDSSEVYSWLD SFLLLMLG--GIPWQAYFQRLSSSSAT-----Y 269
      | | | : | | | || | | :| | : | | |
Db    256 MSPGGLV--SDPISALSLGLALMFGTAGLP--HIIMRFFTVSDAKEARKSVFYATGFIGY 311

Qy    270 AQVLSFLAAFGCLVMAIPAILI----GAIGASTDWNQTAYGLPDPKTTEEADMILPIVLQ 325
      :|:|: || : : | | ||: | | | | |
Db    312 FYILTFIIGFGAILLVSPNPLFKDAAGALIGGT--NMAAVHLAD----- 353

Qy    326 YLCPVYISFFGLGAVS----AAVMSSADSSILSASSMFARNIYQLSFRQ-NASDKEIVWV 380
      | :|| || :| | : : | : :| : :| : :| : | : : |
Db    354 ---AVGGNFF-LGFISAVAFATILAVVAGLTLGASAVSHDLYANVIKNGQADERQELKV 409

Qy    381 MRITVFVFGASATAMALL--TKTVYGLWYLSSDLVYIVIFPQLLCVLFVKGTNTYGAVAG 438
```

```

      :||| : | | : :| : : : | : : || :| : : || | ||| |
Db      410 SKITVVILGIVAIGLGILFEKQNI AFMVGLAFSIAASCNFP IILLSMYWKGLTTRGAVIG 469

Qy      439 YVSGFLRITGGEPYLYLQPLIFYPGYYPDDNGIYNQKFPFKTLAMVTSFLTNICISYLA 498
      ||| : :| : | | | : | || : :| | : :| : :| : :|
Db      470 GWSGLIVAVT----LMILGPTI-WVSILGHDTPIYPYEP-----ALFSMIIAFIV 515

Qy      499 KYLF 502
      :||
Db      516 SWLF 519

```

RESULT 11

US-09-657-960-3

; Sequence 3, Application US/09657960

; Patent No. 6649342

; GENERAL INFORMATION:

; APPLICANT: Mack, David

; APPLICANT: Gish, Kurt

; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSING BREAST CANCER, COMPOSITIONS,  
AND METHODS OF

; TITLE OF INVENTION: SCREENING FOR BREAST CANCER MODULATORS

; FILE REFERENCE: A-69196/DJB/JJD

; CURRENT APPLICATION NUMBER: US/09/657,960

; CURRENT FILING DATE: 2000-09-08

; PRIOR APPLICATION NUMBER: US 09/525,361

; PRIOR FILING DATE: 2000-03-15

; PRIOR APPLICATION NUMBER: US 09/453,137

; PRIOR FILING DATE: 1999-12-02

; PRIOR APPLICATION NUMBER: US 09/450,810

; PRIOR FILING DATE: 1999-11-29

; PRIOR APPLICATION NUMBER: US 09/268,865

; PRIOR FILING DATE: 1999-03-15

; PRIOR APPLICATION NUMBER: PCT/US 00/06952

; PRIOR FILING DATE: 2000-03-15

; NUMBER OF SEQ ID NOS: 4

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 3

; LENGTH: 718

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-657-960-3

Query Match 9.2%; Score 272.5; DB 4; Length 718;

Best Local Similarity 22.3%; Pred. No. 1.2e-17;

Matches 152; Conservative 113; Mismatches 241; Indels 177; Gaps 30;

```

Qy      9 IAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGG- 66
      |||: ::::|:: :| :| :| :| :| :| :| :| :| :|
Db      10 IAIVALYFILVMCIGFFAMWKSNRSTVS-----GYFLAGRSMTWVTIGASL 55

Qy      67 YINGTAEAVYVPGYGLAWAQAPIGYS-----LSLILGGLFFAKPMRSKGYVTMLD 116
      ::: : : ||| : | |:: :| :| :| :| :| :| :|
Db      56 FVSNIGSEHFI---GLAGSGAASGFAVGAWEFNALLLQLLGWVFIPIYIRS-GVYTM-- 109

Qy      117 PFQQIYGKRMGG-----LLFIPALMGEMFWAAAI FSALGATISVIIDVDMHIS 164
      : || || :|:| : : :|:| :| :| :| :| :| :|

```

Db 110 --PEYLSKRFGGHRIQVYFAALSLILYIFTKLSVDLYSGALF-----IQESLGWNLYVS 161  
 QY 165 VIISALIATLYTLVGGLYSVAYTDVVQLFCIFVGLWISVPPFALSHPAVADI-GFTAVHAK 223  
 ||: : | |: || |: || |: : : | : : : | | | :  
 Db 162 VILLIGMTALLTVTGGLVAVIYTDTLQALLMIIG-----ALTLMIISIMEIGGFEEVKRR 216  
 QY 224 YQKPWLGTVDSSSEVYSWLDSF-----LLMLGG-----IPW----- 254  
 | : | : | | : : | | : : | |  
 Db 217 YM-----LASPDVTSILLTYNLSNTNSCNVSPKKEALKMLRNPTDEDVWPWPGFILQOTP 270  
 QY 255 -----QAYFQRLSSSSATYAQ----VLSFLAAFGCLVMAIPAIL-----IGA 293  
 | |||: : : : : | : : : : | : : : : |  
 Db 271 ASVWYWCADQVIVQRLAAKNIAHAKGSTLMAGFLKLLPMFIIVVPGMISRILEFTDDIAC 330  
 QY 294 I-----GASTDWNQTAYGLPDKTTEEADMILPIVLQYLCVPYISFFGLGAVSAAV 344  
 | : : | | : : : | | : : : | | : : : | | :  
 Db 331 INPEHCMLVCGSRAGCSNIAY-----PRLVMKLVVPVGLRGLMMAVMIAAL 375  
 QY 345 MSSADSSILSASSMFARNIYQLSFRQNASDKEIVVMRITV-FVFGASATAMALLTKTVY 403  
 || || |||: : : : : | : : : : | | | : | : : : :  
 Db 376 MSDLDSIFNSASTIFTLDVYKL-IRKSASSRELMIVGRIFVAFMVVISIAWVPIIVEMQG 434  
 QY 404 GLWYLSSDLVYIVIFPQL----LCVLFVKGTNT----YGAVAGYVSGFLRITGGEPYLY 455  
 | || | : | : | : | | | : : : | | | : |  
 Db 435 GQMYLYIQEVADYLTTPVAALFLLAIFWKRCNEQGAFYGGMAGFVLGAVRLILA---FAY 491  
 QY 456 LQPLIFY----PGYYPDDNGIYNQKFPKTLAMVT---SFLT-----NICISYLAKY 500  
 | ||: | : : | : : | | : : | | : : |  
 Db 492 RAPECDQPDNRPGFIKDIHYMYVATGLFWVTGLITVIVSLTTPPTKEQIRTTTFWSKKN 551  
 QY 501 LFESGTLPPKLDVF---DAVVARHSEENMDKTILVKN----ENIK----LDELALVKPRQ 549  
 | || : : : : | | | : : | : : | | | : :  
 Db 552 LVVKENCSPKEEYQMQEKSILRCSENNETINHIIPNGKSEDSIKGLQPEDVNLLVTCRE 611  
 QY 550 SMTLSSTFTNKEAFLDVDSSPEG 572  
 : : : | | : |  
 Db 612 EGNPVASLGHSEAETPVDAYSNG 634

RESULT 12

US-09-134-001C-4744

; Sequence 4744, Application US/09134001C

; Patent No. 6380370

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: GTC-007

; CURRENT APPLICATION NUMBER: US/09/134,001C

; CURRENT FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 60/064,964

; PRIOR FILING DATE: 1997-11-08

; PRIOR APPLICATION NUMBER: US 60/055,779

; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5674

; SEQ ID NO 4744

; LENGTH: 518  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-4744

Query Match 8.8%; Score 262.5; DB 4; Length 518;  
Best Local Similarity 22.2%; Pred. No. 6.9e-17;  
Matches 126; Conservative 102; Mismatches 223; Indels 117; Gaps 25;

```
Qy      9 IAIIVFYLLILLVGIWAARWTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGYI 68
      : |||:::|::| : :| :| :||| || : : |::| |
Db     27 VMIIVFYIILLIIGFY-GYRQATGNLSE-----FMLGGRSIGPYITALSAGASDMSGWMI 80

Qy     69 NGTAEAVYVPGYGLAWAQAPIGYSLSLILGGL--FFAKPMRSKGY-----VTMLDPFQ 119
      | :|| | | :| || :| | :| :| :| :| :|
Db     81 MGLPGSVYSTGLSAIW-----ITIGLTLGAYINYFVAPRLRVYTEIAGDAITLPDFFK 134

Qy    120 QIYGKRMGGLLFIPALMGEMFWAAAFSAL---GATISVIIDVDMHISVIISALIATLYT 176
      : : | | : :| | | :| :| :| :| :| :|
Db    135 NRLDDKKNIKIISGLIIVFFFTLYTHSGFVSGGKLFESAFGLNYHAGLLIVAIIVIFYT 194

Qy    177 LVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADI-GFTAVHAKYQ-KPW----- 228
      || :|: || | : : : : || | : : | : | ||
Db    195 FFGGYLAVSITDFFQGVIMLIAM-VMVPVIV----ALLKLNGWDTFHDIAQMKTPTNLDLFR 249

Qy    229 ----LGTVDSSEVYSWLDSFLLMLGGIPWQAYFQVLSSSSATYAQVLSFLAAFGLVLM 284
      || | :|| :| : :| :| :| :| :| :|
Db    250 GTTVLGIV---SLFSW-----GLGYFGQPHIIVRFMSIKSHKLLPKARRLGISWM 296

Qy    285 AIPAILIGAIGASTDW---NQYAYGLPDKTTEADMILPIVLQYLCPVYISFFGLGAV 340
      |: |::||| : : | ||:| : : : | | : | | :|
Db    297 AVG--LLGAIGVGLTGISFISERHIKLEDPET-----LFIVMSQILFHPLVGGFLLAAI 348

Qy    341 SAAVMSSADSSILSASSMFARNIYQL---SFRQNASDKEIVWVMRITVFVFGASATAMAL 397
      ||::||: | :| || : | :| | : : : || | : :| :| :|
Db    349 LAAIMSTISSQLLVTSSSLTEDFYKLIRGSDKASSHQKEFVLIGRLSVLLVAIVAITIA- 407

Qy    398 LTKTVYGLWYLSSDLVYIV-----IFPQLLCVLFVKGNTNTYGAVAGYVSGFLFRI 447
      |: : : : :| | :| | :| |::| | :| :|
Db    408 -----WHPNDTILNLVGNWAGFGAAFSPLVLYSLYWKDLTRAGAISGMVAGAVVVI 459

Qy    448 TGGEPLYLYLQPLIFYPGYPPDDNGIYNQKFPFKTLAMVTSFLTNICISYLAKYLFESGTL 507
      ::|| : : | :| : : | : : : | :| | |
Db    460 VW---ISWIKPLATINAFF----GMYE-----IIPGFIVSVLITYIVSKL----TK 499

Qy    508 PPKLDVFDVAVVARHSEENMDKTILVKNE 535
      | | : | : :| | |
Db    500 KPD----DYVI-----ENLNKVKHVVKKE 518
```

RESULT 13  
US-09-328-352-6371  
; Sequence 6371, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.

```
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 6371
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-6371
```

```
Query Match          8.7%; Score 259; DB 4; Length 501;
Best Local Similarity 23.1%; Pred. No. 1.4e-16;
Matches 126; Conservative 97; Mismatches 206; Indels 116; Gaps 22;
```

```
Qy      1 MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTA 60
      |::  ||  :::  :::::  |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      6 MSYFDPTLIMFMVYIVAMVLIGLFAYRATSDFSD-----YILGGRSLGSEFVTALSAGA 58

Qy     61 TWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGIFFAKPMR-----SKGYVTML 115
      :  :  |  :  |  |::  |  ||  ||  :  |  |  |  :  |  :  :  :
Db     59 SDMSGWLLMGLPGAIYLSGLSEAW--IAIGLIIGAWLNWLLVAGRLRVHTEIQNNALTLP 116

Qy    116 DPFQQIYGKRMGGLLFIPALMGEMFWAAAIIFSALGATISV-----IIDVDMHISVIISAL 170
      |  |  :  :  |  |::  :::  ||:  |  |  :  :  :  :  :  :  :
Db    117 DYFTSRFDDQKKILRIFSAVILVFF--AIYCASGMVAGARLFENLFGMSYTTAIWLSAI 174

Qy    171 IATLYTLVGGLYSVAYTDVVQLFCIFVGLWISVPFAL-----SHPAVADIGFTAVHAKY 224
      |  :||  ::::||  |  ||  ||  ||  :  :  |  |  :  :  :
Db    175 ATISYVCIGGFLAISWTDTFQ-----AGLMI---FALLLTPIVITYLAIGDTTQFVTLIET 226

Qy    225 QKPWLGTVDSSSEVYSWLDSELLLMLGGIPW-QAYF--QRVLSSSSATYAQVLSFLAAGFC 281
      :|  :  |  |  :  :  |  :  |  ||  :  |  :  :  :  :  :
Db    227 ARPHAFNIIS-----DLSVVAVLSSMAWGLGYFGQPHIL-----VRFMAADS- 268

Qy    282 LVMAIPA-----ILIGAIGASTDWNQTAYGLPDPK----TTEEADMILPIVLQY 326
      |  :||  |  :||:  ||  :  ||  |  :  :  :  :  :  :
Db    269 -VKSIPAARRIGMTWMILCLVGAVGAG--FFGIAYFQQHPELAGVVSKNPETVFMELTKI 325

Qy    327 LCPVYISFFGLGAVSAAVMSSADSSILSASSMFARNIYQLSFRQNASDKEIVWVMRITVF 386
      |  :|  |  |  :  ||||:  :|  ||  :  :  :  |  :||  ||:  ||  ||  |
Db    326 LFNPWIVGIIILAAILAAMSTLSCQLLVCSALTEDLYKGFIRKNASQKELVWVGRIMVL 385

Qy    387 VFGASATAMALLTKTVYGLWYLSSDLVYIVIF-----PQLLCVLFVKGTNTYGAV 436
      |  :|  :  |  |  :  :  |  :  :  :  :  |  :  ||  |  ||:
Db    386 -----AIAVLAIVLAG--NPDSKVLGLVAYAWAGFGAAFGPLIILSLFWKRMTLEGAL 436

Qy    437 AGYVSGFLFRITGGEPYLYLQPLIF--YPGYYPDDNGIYNQKFPFKTLAMVTSFLTNICI 494
      ||  :  |  :  |  |  |  :  :  ||:  :  :  :  :  :  :  :
Db    437 AGMIVGAVVVI--GWKNLFASTGVYEIIPGF-----ICAFISIIVV 475

Qy    495 SYLAK 499
      |  :  :
Db    476 SLISK 480
```

RESULT 14

US-09-252-991A-27829

; Sequence 27829, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
 PSEUDOMONAS  
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 27829  
 ; LENGTH: 551  
 ; TYPE: PRT  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-27829

Query Match 8.7%; Score 257.5; DB 4; Length 551;  
 Best Local Similarity 22.3%; Pred. No. 2.3e-16;  
 Matches 102; Conservative 91; Mismatches 205; Indels 59; Gaps 16;

Qy 6 EGLIAIIVFYLLILLVGI-----WAAWRTKNSGSAEERSEAIIVGGR-DIGLLVGGFTM 58  
 :| |::| |::|: : | | |::|: : | | |::|: : | |  
 Db 28 KGARAMLLDYLIIMLVYALAMLGLGWYGM-----KAKSQSDFLVAGRRLGPGLYLG--TM 80  
 Qy 59 TATWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPF 118  
 | :|| | | : | | | :| |::| |::| : |  
 Db 81 AAVVLGGASTIGTVKLGQFGLSGLWLVFEMLG--LGIIVLSLVFSRQIARLRVFTVTQVL 138  
 Qy 119 QQIY---GKRMGGLLFIPALMGEMFWAAAFSALGATISVIIDVDMHISVIISALIATLY 175  
 :| | : :||:: : : : | |::| : : : : | :|  
 Db 139 EQRYQASSRLIGGVVMVAY---DLMVAVTATIAIGSVTEVVFVIPRIIPAILCGGGIVIVY 195  
 Qy 176 TLVGGLYSVAYTDVVQLFCIFVGLW-ISVPFALSHPA-----VADIGFTAVHAKYQKP 227  
 ::||::|: |::| |::| :| : : | |  
 Db 196 SVIGGMWSLTLTLDIIQFVIKTVGIFLVLLPLSIDGAGGLARMQEVLPAGFFD----- 247  
 Qy 228 WLGTVDSSSEVYSWLDLSFLLMLGGIPWQAYFQRVLSSSSATYAQVLSFLAAFGLVMAIP 287  
 || : : : : || | : | :|| : : | | : | :|  
 Db 248 -LGHIGLDTILTY---FLLYFFGALIGQDIWQRVFTARSETVVRYAGLGAGVYCMLYGAA 303  
 Qy 288 AILIGAIGASTDWNQTAYGLPDPKTTTEADMLPIVLQYLCVPYISFFGL--GAVSAAVM 345  
 |||| | | | : | | | | :|:|  
 Db 304 CALIGAAARVL-----LPDLAVPEN--YAEITREVLAP---GLRGLVVAALSAIM 350  
 Qy 346 SSADSSILSASSMFARNIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKTVYGL 405  
 |:| :|:|:: :|| | : : :| :| :| | |  
 Db 351 STASGCLLAAATVLQEDIYARFLRPGTTS--IRLSRCITLLMGVAMLVLACLVNDVIAA 408  
 Qy 406 WYLSSDLVYIVIFPQLLCVLFVKGTNTYGAVAGYVSG 442

Db                   :: ::|:   :   : | : : ||:| |::|  
409 LSIAYNLLVGGLLVPIVGALLWRRASPQGAIASIVAG 445

RESULT 15

US-09-489-039A-7541  
; Sequence 7541, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
KLEBSIELLA  
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 7541  
; LENGTH: 508  
; TYPE: PRT  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-7541

Query Match                   8.6%;   Score 255;   DB 4;   Length 508;  
Best Local Similarity   24.8%;   Pred. No. 3.6e-16;  
Matches 126;   Conservative   85;   Mismatches 219;   Indels   78;   Gaps   21;

Qy	1	MAFHVEGLIATIVFYLLILLVGIWAAWR-TKNSGSAERSEAIIVGGRDIGLLVGFTMT	59
		:   :   :::  :                 :  :  :  :   :   :	
Db	7	MAISTPMLVTFIVYIFGMVLIG-FIAWRSTKN-----FDDYILGGRSLGPFVTALSAG	58
Qy	60	ATWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMR-----SKGYVTM	114
		: :   :    ::  :        :  :     :  :   :  ::	
Db	59	ASDSMGWLLMGLPGAIFLSGISESW--IAIGLTGAWINWKL VAGRLRVHTEVNNNALTL	116
Qy	115	LDPFQQIYGKRMGGLLFIPALMGEMFWAAAI FSALGATISV-----IIDVDMHISVIISA	169
		: :       : :: :               :   ::	
Db	117	PDYFTGRFEDKSRVLRISALVILLFF--TIYCASGIVAGARLFESTFGMSYETALWAGA	174
Qy	170	LIATLYTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADI---GF-TAVHAKYQ	225
		:       : ::      :     ::     :      ::	
Db	175	AATIIYTFVGGFLAVSWTDTVQASLMIFALILT-----PVIVII SVGGFGDSLEVIKQ	227
Qy	226	KPWLGTVDSSSEVYSWLDSFLLMLGGIPW-----QAYFQRVLSSSSATYAQVLSF	275
		::: ::   : : :                         : : :	
Db	228	K----SIENIDMLKGLNFVAIISIMG--WGLGYFGQPHILARFMAADSHHSIVHARRISM	281
Qy	276	LAAFGCLVMAIPAILIGAIGASTDWNQTAYGLPDPK----TTEEADMILPIVLQYLCVPY	331
		: :         :    :  : : :     :   :	
Db	282	TWMILCLG---GAVAVGFFG-----IAYFNNNP SLAGAVNQNAERVFIELAQILFNPW	331
Qy	332	ISFFGLGAVSAAVMSSADSSILSASSMFARNIYQLSFRQNASDKEIVWVMRITVFVFGAS	391
		:    :     : :      :::  ::    :     :	
Db	332	IAGILLSAILAAVMSTLSCQLLVCSAITEDLYKAFLRKNAGQKELVWVGRMMVLVALV	391

Qy 392 ATAMA-----LLTKTVYGLWYLSSDLVYIVIFPQLLCVLFVKGTNTYGAVAGYVSGFL 445  
 | |:| :| | : |:| |: : | |:| | |  
 Db 392 AIALAANPENRVLGLVSYAWAGFGAAGPVLV----SVMWSRMTRN-GALAGMVIGALT 446  
 Qy 446 RITGGE-PYLYLQPLIFYPGYYPDDNGI 472  
 | : :| | :| ||: ||  
 Db 447 VIVWKQFGWLGLYEII--PGFVFGSIGI 472

Search completed: March 22, 2004, 15:36:40  
 Job time : 39 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 22, 2004, 15:19:38 ; Search time 40 Seconds  
(without alignments)  
1394.777 Million cell updates/sec

Title: US-10-069-541-6  
Perfect score: 2972  
Sequence: 1 MAFHVEGLIAIIVFYLLILL.....EAFLDVDSSPEGSGTEDNLQ 580

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_78:\*  
1: pirl:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	2972	100.0	580	2	JC7502	choline transporte
2	1361.5	45.8	631	2	T20037	hypothetical prote
3	344	11.6	492	2	D75188	proline symporter
4	308.5	10.4	662	2	A37226	glucose transport
5	306	10.3	664	2	A33545	Na+/glucose cotran
6	306	10.3	665	2	A53582	Na+/glucose cotran
7	303.5	10.2	660	2	A44432	amino acid transpo
8	301	10.1	463	2	E83468	probable sodium/so
9	299.5	10.1	507	2	B83988	proline transporte
10	298	10.0	672	2	A42251	nucleoside transpo
11	294	9.9	664	2	S59637	glucose transport
12	292	9.8	672	2	A56765	sodium-glucose cot
13	288	9.7	664	2	S59638	glucose transport

14	286	9.6	491	2	H71097	hypothetical prote
15	285	9.6	492	2	H69670	sodium/proline sym
16	284	9.6	537	2	C71008	probable proline p
17	283	9.5	484	2	E75138	osmoregulated prol
18	282.5	9.5	501	2	T44298	sodium/proline sym
19	279.5	9.4	496	2	H82382	sodium/proline sym
20	278.5	9.4	718	2	A42163	Na+/myo-inositol c
21	271	9.1	537	2	A75123	proline permease (
22	270.5	9.1	718	2	A56851	Na+/myo-inositol c
23	267	9.0	522	2	B84247	proline permease [
24	263.5	8.9	494	2	JC2382	sodium/proline sym
25	261	8.8	512	2	E89978	high affinity prol
26	260.5	8.8	504	2	E64118	sodium/proline sym
27	260.5	8.8	605	2	A36361	glucose transport
28	259	8.7	526	2	C69115	sodium/proline sym
29	254	8.5	496	2	A71980	sodium/proline sym
30	254	8.5	502	2	E90786	major sodium/proli
31	254	8.5	502	2	C85646	major sodium/proli
32	253.5	8.5	497	2	T48676	proline uptake pro
33	253	8.5	502	1	JGECPP	sodium/proline sym
34	252	8.5	461	2	G83610	probable sodium/so
35	249	8.4	496	2	G64526	sodium/proline sym
36	249	8.4	502	2	AF0633	sodium/proline sym
37	246.5	8.3	512	2	S75887	hypothetical prote
38	246	8.3	502	2	S10220	sodium/proline sym
39	243.5	8.2	506	2	F83547	sodium/proline sym
40	243	8.2	497	2	AH2015	sodium/solute symp
41	242	8.1	449	2	B69759	sodium/proline sym
42	239.5	8.1	484	2	H64105	pantothenate trans
43	239	8.0	549	2	B65215	hypothetical 59.2
44	237.5	8.0	577	2	T28017	hypothetical prote
45	236	7.9	491	2	E69383	pantothenate perme

#### ALIGNMENTS

##### RESULT 1

JC7502

choline transporter - human

C;Species: Homo sapiens (man)

C;Date: 17-Nov-2000 #sequence\_revision 17-Nov-2000 #text\_change 01-Dec-2000

C;Accession: JC7502

R;Apparsundaram, S.; Ferguson, S.M.; George Jr., A.L.; Blakely, R.D.

Biochem. Biophys. Res. Commun. 276, 862-867, 2000

A;Title: Molecular cloning of a human, hemicholinium-3-sensitive choline transporter.

A;Reference number: JC7502

A;Contents: Spinal cord

A;Accession: JC7502

A;Molecule type: mRNA

A;Residues: 1-580 <APP>

A;Cross-references: GB:AF276871

C;Comment: This protein, a hemicholinium-3-sensitive phosphorylated transmembrane protein, regulates high-affinity choline uptake, and plays the roles in disease states.

C;Genetics:

A;Gene: cht

A;Map position: 2q12

C;Keywords: choline transport; spinal cord; transmembrane protein; transport protein

Query Match 100.0%; Score 2972; DB 2; Length 580;  
Best Local Similarity 100.0%; Pred. No. 9.5e-211;  
Matches 580; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY      1 MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTA 60
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTA 60

QY     61 TWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ 120
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ 120

QY    121 IYGKRMGGLLFIPALMGEMFWAAAI FSALGATISVIIDVDMHISVIISALIATLYTLVGG 180
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 IYGKRMGGLLFIPALMGEMFWAAAI FSALGATISVIIDVDMHISVIISALIATLYTLVGG 180

QY    181 LYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSEVYSW 240
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 LYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSEVYSW 240

QY    241 LDSFLLMLGGIPWQAYFQVLSSSSATYAQVLSFLAAGCLVMAIPAILIGAIGASTDW 300
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 LDSFLLMLGGIPWQAYFQVLSSSSATYAQVLSFLAAGCLVMAIPAILIGAIGASTDW 300

QY    301 NQTAYGLPDPKTTEEADMILPIVLQYLCVPYISFFGLGAVSAAVMSSADSSILSASSMFA 360
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 NQTAYGLPDPKTTEEADMILPIVLQYLCVPYISFFGLGAVSAAVMSSADSSILSASSMFA 360

QY    361 RNIYQLSFRQNASDKEIVWMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQ 420
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 RNIYQLSFRQNASDKEIVWMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQ 420

QY    421 LLCVLFVKGTNTYGAVAGYVSGFLRITGGEPYLYLQPLIFYPGYYPDDNGIYNQKFPFK 480
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 LLCVLFVKGTNTYGAVAGYVSGFLRITGGEPYLYLQPLIFYPGYYPDDNGIYNQKFPFK 480

QY    481 TLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDVAVVARHSEENMDKTILVKNENIKLD 540
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDVAVVARHSEENMDKTILVKNENIKLD 540

QY    541 ELALVKPRQSMTLSSTFTNKEAFLDVDSSPEGSGTEDNLQ 580
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 ELALVKPRQSMTLSSTFTNKEAFLDVDSSPEGSGTEDNLQ 580
```

RESULT 2

T20037

hypothetical protein C48D1.3 - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C;Accession: T20037

R;Burton, J.

submitted to the EMBL Data Library, October 1996

A;Reference number: Z19214

A;Accession: T20037

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-631 <WIL>

A;Cross-references: EMBL:Z81049; PIDN:CAB02847.1; GSPDB:GN00022; CESP:C48D1.3

A;Experimental source: clone C48D1

C;Genetics:

A;Gene: CESP:C48D1.3

A;Map position: 4

A;Introns: 82/1; 120/3; 187/1; 236/3; 249/1; 358/1; 510/3; 570/2

Query Match 45.8%; Score 1361.5; DB 2; Length 631;  
Best Local Similarity 46.8%; Pred. No. 2.5e-92;  
Matches 290; Conservative 91; Mismatches 146; Indels 93; Gaps 10;

```
Qy      7 GLIAIIVFYLLILLVGIWAAWRTKNSGSAEER----SEAIIVGGRDIGLLVGGFTMTATW 62
      |::||: ||:||||:|||| |::| | :| :: ||:| || | |||||
Db      6 GIVAIVFFYVLILVVGIWAGRKSKSSKELESEAGAATEEVMLAGRNIGTLVGIFTMTATW 65

Qy     63 VGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDP----- 117
      ||| |||||:| | |||:|::|::|:| || | | :|:| ||||
Db     66 VGGAYINGTAEALY--NGGLLGCQAPVGYAISLVMGGLLFAKKMREEGYITMLDPFQFWN 123

Qy    118 -----FQQIYGKRMGGLLFIPALMGEMFWAAAI F 146
      || ||:|:|::|:|:|:| || || |
Db    124 FLELIFGRTFDNFRKLGRFLKLQTIIEILDFFQHKGQRIGGLMYVPALLGETFWTAAIL 183

Qy    147 SALGATISVIIDVDMHISVIISALIATLYTLVGGLYSVAYTDVVQLFCIFVGLWISVPFA 206
      |||||:|:|: |:: || :|| || || || |::| ||||| || | :
Db    184 SALGATLSVILGIDMNASVTLASACIAVFYFTFTGGYYAVAYTDVVQLFCIFVGLLILGLYV 243

Qy    207 LSHPAVADIGFTAVHAKYQKPWLGTVDSSSEVSWLDSFLLMLGGIPWQAYFQRVLSSSS 266
      : | | | | | | | | | | | | | | | | | | | | | | :
Db    244 QNRPN-----RFKETSLEWIDCMLLVFGGIPWQVYFQRVLSST 282

Qy    267 ATYAQVLSFLAAGCLVMAIPAILIGAIGASTDWNQTAYGLPDKTTEEA-----DMIL 320
      | || |||:| ||:| ||| |||| |::| | | : | | : |::
Db    283 AHGAQTLSEFVAGVGCILMAIPALIGAIARNTDWRMTDYSPPWNGTKVESIPDPKRNMVV 342

Qy    321 PIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFARNIYQLSFRQNASDKIIVWV 380
      |:| || | :::| ||||| |||||:|:|:| ||:|:| | :|:|:|:|
Db    343 PLVFQYLTFRWVAFIPLGAVSAAVMSSADSSVLSAASMFANINWKLITIRPHASEKEVIIV 402

Qy    381 MRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQLLCVLFVKGTNTYGAAGYV 440
      ||| : | || || | ::| |||| |::|:|:|:|:|:|:|:|:|:|
Db    403 MRIAIICVGIMATIMALTQSIYGLWYLCADLVYVILFPQLLCVYMPRSNTYGSAGYA 462

Qy    441 SGLFLRITGGEPYLYLQPLIFYPGYPPDDNGIYNQKFPFKTLAMVTSFLTNICISYLAKY 500
      || ||: ||| : | || | :|: | ||:| ||:| | :| :
Db    463 VGLVLRILIGGEPLVSLPAFFHYPMY---TDGV--QYFPFRTTAMLSSMATIYIVSIQSEK 517

Qy    501 LFESGTLPPKLDVFDVAVVARHSEENMDKTILVKNENIKLDELALVKPRQSMTLSSTFTNK 560
      ||:| | |: || || | | : | : | | :| | |
Db    518 LFKSGRLSPEWDMGCVV-----NIPIDHVPLPSD-VSFAVSSETLNM 559
```

Qy 561 EAFLDVDSPEGSGTEDNLQ 580  
: | : | : || ||  
Db 560 KVECDGMQFPQ-LQTEHRLQ 578

RESULT 3

D75188

proline symporter (proline permease). PAB2354 - Pyrococcus abyssi (strain Orsay)

C;Species: Pyrococcus abyssi

C;Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000

C;Accession: D75188

R;anonymous, Genoscope

submitted to the EMBL Data Library, July 1999

A;Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure and evolution.

A;Reference number: A75001

A;Accession: D75188

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-492 <KAW>

A;Cross-references: GB:AJ248283; GB:AL096836; NID:g5457433; PIDN:CAB48955.1; PID:g5457464

A;Experimental source: strain Orsay

C;Genetics:

A;Gene: putP-3; PAB2354

C;Superfamily: proline carrier protein

Query Match 11.6%; Score 344; DB 2; Length 492;  
Best Local Similarity 24.2%; Pred. No. 1.1e-17;  
Matches 132; Conservative 99; Mismatches 196; Indels 118; Gaps 25;

Qy 8 LIAIIVFYLLILLVGIWAARWTKNSGSAERSEAIIVGGRDIGLLVGGFTMTATWVG GGY 67  
|:| : :| :||| :| ||| : : : :  
Db 14 LVAFLETLILPILVGFYAMKRTKS-----EEDFFVGGGRAMDKITVALSAVSSGRSSWL 66

Qy 68 INGTAEEAVYVPGYGLAWAQAPIGYSLs----LILGGLFFAKPMRSKGYVTMLDPFQQIYG 123  
: | : | | | | :|| : : | : | : | :  
Db 67 VLGLSGMAYKMGVTA VV--AAVG YIVAEMFQFVYMGIRLRKFSE RFNAITVPDYFEARFR 124

Qy 124 K-----RMGG----LLFIPALMGEMFWAAAFSALGATISVIIDVDMHISVIISALIATL 174  
| : : : :| :| | | | :| : : : :|| | : :  
Db 125 DTSKILRIAASIIIIIFLTSYVGAQFNAGA-----KTLSTALGISIFTALMISVLMIIIV 178

Qy 175 YTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFT-----AVHAKYQK 226  
| :|| :||| ||: : :|| : | || :| | : |  
Db 179 YMILGGFIAYAYNDVIRAVIMIIGLVV-----LPVIAVAKVGGTEEV LKVLHALDPK LIN 233

Qy 227 PW---LGTVDSSSEVYSWLDSFLLMLG-GIPWQAY-FQVLSSSSATYAQVLSFLAAFGC 281  
|| | | | :| | | : | :| : : : |  
Db 234 PWAFGAGVVIG-----FLGIGFGSPGQPHIIVRYMSIDDPNKL RVSTVVGTFWN 282

Qy 282 LVMAIPAILIGAIGASTDWNQTAYGLPDPKTT--EEADMILP-IVLQYLCPVYISFFGLG 338  
:|:| || :| | : :|| : :|| | : || | : :|  
Db 283 VVLAWGAIFVGLAGRAI-----VPDVSQ L PGKNAEMIYPYLSAQYFPPIY GIL-IG 333

Qy 339 AVSAAMVSSADSSILSASSMFARNIYQLSFRQNA--SDKEIVWMRITVFVFGASATAMA 396  
: ||::|:| | :| :| :|| : : :| :| | | | :|

Db 334 GIFAAILSTADSQLLVVASTVVVKDLYQEVIKKGTKIDEKTALTISRVTVLVVGFLAAILA 393

Qy 397 LLTKTVYGLWYLSSDLVY-IVIF-----PQLLCVLFVKGTNTYGA VAGYVSGLFL 445  
 |::|:::|:| |::|:|:|:|:|:|:|

Db 394 -----YVAKDIIFFWFLFAWGGLGASFGPTLILSLYWKGTTKWGVLAGMIVGTIT 443

Qy 446 RITGGEPYLYLQPLIFYPGYPPDDNGIYNQKFPFKTLAMVTSFLTNICISYLAKYLFESG 505  
 | |::|: |::|:|:|:|:|:|:|:|

Db 444 TIVW---KLYLKPI-----TGLY-ELVP----AFIFSLIATIIVSMITK----- 479

Qy 506 TLPPK 510  
 |:

Db 480 --PPE 482

#### RESULT 4

A37226

glucose transport protein - rabbit

N;Alternate names: sodium/D-glucose cotransporter

C;Species: *Oryctolagus cuniculus* (domestic rabbit)

C;Date: 30-Dec-1991 #sequence\_revision 01-Mar-1996 #text\_change 20-Aug-1999

C;Accession: S00515; S15974; A37226

R;Hediger, M.A.; Coady, M.J.; Ikeda, T.S.; Wright, E.M.

Nature 330, 379-381, 1987

A;Title: Expression cloning and cDNA sequencing of the Na/glucose co-transporter.

A;Reference number: S00515; MUID:88065856; PMID:2446136

A;Accession: S00515

A;Molecule type: mRNA

A;Residues: 1-662 <HED>

A;Cross-references: EMBL:X06419; NID:g1640; PIDN:CAA29727.1; PID:g1641

R;Morrison, A.I.; Panayotova-Heiermann, M.; Feigl, G.; Schoelermann, B.; Kinne, R.K.H.

Biochim. Biophys. Acta 1089, 121-123, 1991

A;Title: Sequence comparison of the sodium-D-glucose cotransport systems in rabbit renal and intestinal epithelia.

A;Reference number: S15974; MUID:91223090; PMID:2025641

A;Accession: S15974

A;Molecule type: mRNA

A;Residues: 1-662 <MOR>

A;Cross-references: EMBL:X55355; NID:g1716; PIDN:CAA39040.1; PID:g1717

R;Coady, M.J.; Pajor, A.M.; Wright, E.M.

Am. J. Physiol. 259, C605-C610, 1990

A;Title: Sequence homologies among intestinal and renal Na(+)/glucose cotransporters.

A;Reference number: A37226; MUID:91023017; PMID:2221040

A;Accession: A37226

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 178-662 <COA>

A;Cross-references: GB:X06419

A;Experimental source: renal cortex

C;Superfamily: proline carrier protein

Query Match 10.4%; Score 308.5; DB 2; Length 662;

Best Local Similarity 23.4%; Pred. No. 6.6e-15;

Matches 154; Conservative 110; Mismatches 238; Indels 155; Gaps 26;

Qy 11 IIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGYING 70  
|:::|::: ||:| | : | | : : || : | :: | : | : |  
Db 32 IVIYFLVVMVAVGLWAMFST-NRGTV----GGFFLAGRSMVWVPIGASLFASNIGSGHFGV 86

Qy 71 TAEAVYVPGYGLAWAQAPIGYS-----LSLILGGLFFAKPMRSKGYVTMLDPFQQIY-GK 124  
| | | | | : : : | : | : | | | : | : : |  
Db 87 LA-----GTGAASGIATGGFEWNALIMVVVLGWVFPYIYIRA-GVVTMPEYLQKRFGGK 139

Qy 125 RMGGLLFIPALMGEMFW--AAAFSALGAT-ISVIIDVDMHISVIISALIATLYTLVGGL 181  
| : | | : : : | : | | | | : : : : : : | : | | | : | |  
Db 140 RIQIYLSILSLLLYIFTKISADIFS--GAIFIQLTLGLDIYVAIIILLVITGLYITITGGL 197

Qy 182 YSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKY-----Q 225  
: | | | : | : | | | | : | | | | : | |  
Db 198 AAVIYTDLTQTAIMMVGSVILTGFATHEVG---GYEAFTEKYMRAIPSQISYGNSTSIPO 253

Qy 226 KPWLGTVDSSSEVYSWLDLSFLLMLGGIPW-----QAYFQVRVLSSSSA 267  
| : | : : | : | | | | | | | | : :  
Db 254 KCYTPREDAFHI-----FRDAITGDIPWPLVFGMSILTLWYWCTDQVIVQRCLSAKNL 307

Qy 268 TYAQVLSFLAAGCLVMAIPAILIGAIGASTDWNQTAYGLPDP-----KTTEEADMILP 321  
: : : | : : : : : | : : | : : : : |  
Db 308 SHVKAGCILCGYLKVMFLVMMGMVSRILYTDKVCVVPSECERYCGTRVGCTNIAFP 367

Qy 322 IVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFARNIYQLSFRQNASDKEIVWVM 381  
: : | | : | : : : | | | | : | | : | : : : :  
Db 368 TLVVELMPNGLRGLMLSVMMASLMSSTLSIFNSASTLFTMDIY-TKIRKKASEKELMIAG 426

Qy 382 RI-TVFVFGASATAMALLTKTVYG--LWYLSSDLVYI--VIFPQLLCVLFVKGTNTYGA 436  
| : : : | | : : : | : | : | : | : | | | |  
Db 427 RLEMLFLIGISIAWVPIVQSAQSGQLFDYIQSITSYLGPPIAAVFLLAIFWKRVNEPGAF 486

Qy 437 AGYVSGLEFLRI-----TG-----GEPYLYLQPLIFYPGYPPDDNGIY 473  
| | | : | | | | | | | | : : |  
Db 487 WGLVLGFLIGISRMITEFAYGTGSCMEPSNCPTIICGVHYLYFAIILF----- 534

Qy 474 NQKFPFKTLAMVTSFSLTNICISYLAKYLFESGTLPPKLDVFDVAVVA-RHSEENMDKTILV 532  
| | : | : : | : | : : : : : : :  
Db 535 -----VISIITVVVVSFLTKEPI-----PDVHLYRLCWSLRNSKE----- 568

Qy 533 KNENIKLD--ELALVKPRQSMTSLSTFTNKEAF-----LDVDSSPEGSGTED 577  
| | | | : : : : : | : | | | : : | :  
Db 569 --ERIDLDAGEEDIQEAPPEATDTEVPKKKKGFFRRAYDLFCGLDQDKGPKMTKEEE 623

RESULT 5

A33545

Na<sup>+</sup>/glucose cotransporter SGLT1 - human

C;Species: Homo sapiens (man)

C;Date: 27-Feb-1990 #sequence\_revision 27-Feb-1990 #text\_change 20-Aug-1999

C;Accession: A33545; A53804

R;Hediger, M.A.; Turk, E.; Wright, E.M.

Proc. Natl. Acad. Sci. U.S.A. 86, 5748-5752, 1989

A;Title: Homology of the human intestinal Na<sup>(+)</sup>/glucose and Escherichia coli Na<sup>(+)</sup>/proline cotransporters.

A;Reference number: A33545; MUID:89345544; PMID:2490366

A;Accession: A33545  
 A;Molecule type: mRNA  
 A;Residues: 1-664 <HED>  
 A;Cross-references: GB:M24847; NID:g338054; PIDN:AAA60320.1; PID:g338055  
 R;Turk, E.; Martin, M.G.; Wright, E.M.  
 J. Biol. Chem. 269, 15204-15209, 1994  
 A;Title: Structure of the human Na<sup>+</sup>/glucose cotransporter gene SGLT1.  
 A;Reference number: A53804; MUID:94253082; PMID:8195156  
 A;Accession: A53804  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-45 <TUR>  
 A;Note: sequence extracted from NCBI backbone (NCBIN:147993, NCBIP:147994)  
 C;Genetics:  
 A;Gene: GDB:SLC5A1; SGLT1  
 A;Cross-references: GDB:120375; OMIM:182380  
 A;Map position: 22q13.1-22q13.1  
 C;Superfamily: proline carrier protein  
 C;Keywords: transmembrane protein; transport protein

Query Match 10.3%; Score 306; DB 2; Length 664;  
 Best Local Similarity 22.8%; Pred. No. 1e-14;  
 Matches 148; Conservative 104; Mismatches 218; Indels 178; Gaps 30;

```

Qy      11 IIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGYING 70
        |: ::::: ||:| | : | | :      : || :   | :: |: :| |: |
Db      32 IVIYFVVVMAVGLWAMFST-NRGTV----GGFFLAGRSMVWVPIGASLFASNIGSGHFVG 86

Qy      71 TAEAVYVPGYGLAWAQAPIGYS-----LSLILGGLFFAKPMRSK-GYVTMLDPFQQIYGK 124
        |   | | | | | |:      | ::|| ||   |: | | |||   : |
Db      87 LA-----GTGAASGIAIGGFENALVLVVVLGWLFV--PIYIKAGVVTM----PEYLRK 134

Qy     125 RMGG-----LLFIPALMGEMFWAAAI FSALGATISVIIDVDMHISVIISALIA 172
        | ||      ||:| :   :: |||      |: : ::::: : |
Db     135 RFGGQRIQVYLSLLSLLLYIFTKISADIFSGAIF-----INLALGLNLYLAIFLLLAIT 188

Qy     173 TLYTLVGGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQK--PWL- 229
        |||: ||| :| ||| :| : || |   ||      |: |   || | | :
Db     189 ALYTITGGLAAVIYTDLTQTVMILVGLSLILTGF AFHEVG----GYDAFMEKYMKAIPTIV 244

Qy     230 ---GTVDSSEVYS-WLDSFLLL---MLGGIPW-----QAYFQRVLSS 264
        |   : |: ||| :   : | :||      |   || |||:
Db     245 SDGNTTFQEKCYTPRADSFHIFRDPLTGDLPWPGFIFGMSILTLWYWCTDQVIVQRCLSA 304

Qy     265 SSATY AQ----VLSFLAAFGLVMAIPAIL-----IGAI-----GASTDWNQT 303
        : :: :   : :|   :| :| ::      | :      |
Db     305 KNMSHVKGGCILCGYLKLMPMFIMVMPGMISRILYTEK IACVVPSECEKYCGTKVGCTNI 364

Qy     304 AYGLPDPKTT EADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFARNI 363
        ||      | :: | | :   | : |: ||| |   |||:| :|
Db     365 AY-----PTLVVELMPNGLRGLMLSVMLASLMSSLT SIFNSASTLFTMDI 409

Qy     364 YQLSFRQNASDKEIVWVMRITVFV-FGASATAMALLTKTVYG--LWYLSSDLVYI--VIF 418
        |   |: ||: ||: :   |: : | | | : ::   |   |: |   |: |
Db     410 Y-AKVRKRASEKELMIAGRLFILVLIGISIAWVPIVQSAQSGQLFDYIQSITSYLGPPIA 468

Qy     419 PQLLCVLFVKGTNTY GAVAGYVSGLFLRI-----TG-----GEPYLY 455
  
```



```

      | : | | | | | : | | : | | | | | | | |
Db      469 AVFLLAIFWKRVNEPGAFLWGLILGLLIGISRMITEFAYGTGSCMEPSNCPTIICGVHYLY 528

Qy      456 LQPLIFYPGYYPDDNGIYNQKFPFKTLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFD 515
      : : | | | | | | | | | | | | | | | | | |
Db      529 FAILF-----AISFITIVVISLLTKPI-----PDVHLYR 558

Qy      516 AV--VARHSEENMDKTILVKENIKLDELALVKPRQSMSTLSSTFTNKE 561
      : | | : | : : | | : | : : : : | :
Db      559 LCWSLRNSKEERID--LDAEEENIQ-----EGPKETIEIETQVPEKK 598

```

RESULT 6

A53582

Na<sup>+</sup>/glucose cotransporter SGLT1 - rat

C;Species: Rattus norvegicus (Norway rat)

C;Date: 12-Apr-1995 #sequence\_revision 12-Apr-1995 #text\_change 20-Aug-1999

C;Accession: A53582

R;Lee, W.S.; Kanai, Y.; Wells, R.G.; Hediger, M.A.

J. Biol. Chem. 269, 12032-12039, 1994

A;Title: The high affinity Na<sup>+</sup>/glucose cotransporter. Re-evaluation of function and distribution of expression.

A;Reference number: A53582; MUID:94216314; PMID:8163506

A;Accession: A53582

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-665 <LEE>

A;Cross-references: GB:U03120; NID:g414571; PIDN:AAA19015.1; PID:g414572

C;Superfamily: proline carrier protein

```

Query Match          10.3%; Score 306; DB 2; Length 665;
Best Local Similarity 23.5%; Pred. No. 1e-14;
Matches 155; Conservative 105; Mismatches 242; Indels 158; Gaps 29;

```

```

Qy      11 IIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGYING 70
      | : : : : : | : | : | : | : | : | : | : | : |
Db      32 IVIYFVVVMAVGLWAMFST-NRGTV----GGFFLAGRSMVWWPIGASLFASNIGSGHFGV 86

Qy      71 TAEAVYVPGYGLAWAQAPIGYSL-----LILGGLFFAKPMRSK-GYVTMLDPFQQIYGK 124
      | | | | | | : : : : | : | | | | : |
Db      87 LA-----GTGAAAGIAMGGFEWNALVFVVVLGWLFV--PIYIKAGVVTM----PEYLRK 134

Qy      125 RMGG-----LLFIPALMGEMFWAAAFSALGATISVIIDVDMHISVIISALIA 172
      | | | | | | : : : : | : : : : : | |
Db      135 RFGGKRIQIYLSVLSLLLYIFTKISADIFSGAIF-----INLALGLDIYLAIFILLAIT 188

Qy      173 TLYTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQK--PWL- 229
      | | : | | : | : | : | | | : | | | | |
Db      189 ALYTITGGLAAVIYTDLTQTAIMLVGSFILTGFVAFREVG----GYEAFMDKYMKAIPTLV 244

Qy      230 --GTVD-SSEVYS-WLDSFLLL---MLGGIPW-----QAYFQRVLS 264
      | : | : | : : : | : | | | | | : | | | :
Db      245 SDGNITVKEECYTPRADSFIHFRDPITGDMWPWGLIFGLSILALWYWCTDQVIVQRCLSA 304

Qy      265 SSATYAQVLSFLAAGCLVMAIPAILIGAIGASTDWNQTAYGLPDP-----KTTEEADM 318
      : : : | : : : | | : | | : :
Db      305 KNMSHVKAGCTLCGYLKLPMFLMVMPGMISRILYTDKACVLPSECKKYCGTPVGCTNI 364

```





A;Cross-references: GB:AE004571; GB:AE004091; NID:g9947360; PIDN:AAG04807.1;  
 GSPDB:GN00131; PASP:PA1418  
 A;Experimental source: strain PA01  
 C;Genetics:  
 A;Gene: PA1418

Query Match 10.1%; Score 301; DB 2; Length 463;  
 Best Local Similarity 25.1%; Pred. No. 1.5e-14;  
 Matches 115; Conservative 86; Mismatches 211; Indels 46; Gaps 15;

```

Qy      9 IAIIVFYLLILLVGI----WAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGF---TMTAT 61
      :|: :| :| | | : | | | : | : | | :| | | | |
Db      1 MALDIFVVLIIYAAGMIALGWYGMR-----RAKTRDD-YLVAGRNLG---PGFYLGTMAT 51

Qy     62 WVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQOI 121
      :| | | | | | | | | | | | : | : | | | : | : |
Db     52 VLGGASTIGTVRLGYVHGISGFWLCAIG--LGIVGLSLFLAKPLLKLIYTVTQVLERR 109

Qy    122 YGKRMGGLLFIPALMGEMFWAAAFSALGATISVIIDVDMHISVIISALIATLYTLVGG 181
      | : | : | | | : | : | : | : | : | : | : |
Db    110 YNPAARHASALIMLVYALMIGATSTIAIGTVMQVLFGLPFVWSILIGGGVVVLYSTIGGM 169

Qy    182 YSVAYTDVVQLFCIFVGL-WISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSSEVYSW 240
      :|: ||:| | : ||| : : | : : | : | : | : | : |
Db    170 WSLTLTDIVQFLIMTVGLVFLMLPLSINDAG----GWDALVAKLPASYF---DFTAI-GW 221

Qy    241 ---LDSFLLMLGGIPWQAYFQRVLSSSSATYAQVLSFLAAGCLVMAIPAILIGAIGAS 297
      : | | : | | : ||| : : | | | : | | : : | | |
Db    222 DTIVTYFLIYFFGIFIGQDIWQRVFTARSETVAKVAGSAAGIYCVLYGMAGALIGMAAKV 281

Qy    298 TDWNQTAYGLPDPKTEEADMILPIVLQYLCPVYISFFGLGAVSAAMVSSADSSILSASS 357
      || | | : | : : | | : | | : | : | : | : | : |
Db    282 L-----LPD----LENVNNAFASVVEHSLPNGIRGLVIAAALAALMSTASAGLLAAST 330

Qy    358 MFARNIY-QLSFRQNASDKEIVVWMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIV 416
      : : : | : | | | | | : | : | : | | : : : | :
Db    331 TVTQDLLPRLRRGRGQSDNGDVHENRIATLLLGLVVLGIALVVSDDVISALTVAYNLLVGG 390

Qy    417 IFPQLLCVLFVKGTNTYGAVA----GYVSGFLFRITGG 450
      : | : : | | | : | : : | | |
Db    391 MLIPLIGAIYWKRATTAGAITSMTLGFLTTLVFMKIDG 428
  
```

# RESULT 9

B83988

proline transporter opuE [imported] - *Bacillus halodurans* (strain C-125)

C;Species: *Bacillus halodurans*

C;Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001

C;Accession: B83988

R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hiram, C.; Nakamura, Y.; Ogasawara, N.; Kuhara, S.; Horikoshi, K.

Nucleic Acids Res. 28, 4317-4331, 2000

A;Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and genomic sequence comparison with *Bacillus subtilis*.

A;Reference number: A83650; MUID:20512582; PMID:11058132

A;Accession: B83988

A;Status: preliminary

A;Molecule type: DNA  
A;Residues: 1-507 <STO>  
A;Cross-references: GB:AP001516; GB:BA000004; NID:g10175192; PIDN:BAB06425.1;  
GSPDB:GN00137  
A;Experimental source: strain C-125  
C;Genetics:  
A;Gene: opuE  
C;Superfamily: proline carrier protein

Query Match 10.1%; Score 299.5; DB 2; Length 507;  
Best Local Similarity 26.2%; Pred. No. 2.2e-14;  
Matches 141; Conservative 84; Mismatches 220; Indels 93; Gaps 27;

```

Qy      5 VEGL-IAIIVFYLL-ILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATW 62
      || | :||:: ||: :||:|: :: : : | : : ||:: : ::
Db      4 VEPLAVAILIAYLVALLLIGLLSS-KKSSVGMTD-----FFIAGRNLNKWTVALSAVSSG 57

Qy     63 VGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGY-----VTMLD 116
      : | | | | | | | : || | : | : | : |
Db     58 RSAWLVLGVTGTAYATGLDAVWAVA--GYITVEVF--LFFYVARRFRAYSEQTGSITIPD 113

Qy    117 PFQQIYGKR----MGGLLFIPALMGEMFWAAAFSAL---GATISVIIDVDMHISVIISA 169
      : : : || || :| | | : | | | : : | : ::|
Db    114 ILETRFNDKTHILRGGSAFI--IM--FFMIAYVASQLVAGGGAFATSMGVSSSTGMWVTA 169

Qy    170 LIATLYTLVGGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIG-----FTAVHA 222
      :| ||::|| ::|: |||| :|| | | | || || :|
Db    170 VILLAYTMLGGFHAVSKTDVVQAGFMFVSLVIL-----PVVAIIGLGGFDPLLQVMHT 222

Qy    223 KYQKPWLGTVDSSSEVYSWLDSFLLMLG-GIPWQAY-FQVLSSSSATYAQVLSFLAAF 280
      : | | : : || :| | | : | : | : : ::
Db    223 EG-----GGFTSPFAFGFAGVIGLLGIGFGSPGNPHILVRYMSLKNVKEMRQAALISSVW 277

Qy    281 CLVMAIPAILIGAIGASTDWNQTAYGLPDPKTTEEAD---MILPIVLQYLCVPYISFFGL 337
      ::| |::|| | || || | : | : | |:: | |
Db    278 NVLMGWGAVMIGLAG-----RAY-FPDVSLLPNGDQEQVFLMLGSEILHPLFFGFL-L 328

Qy    338 GAVSAAVMSSADSSILSASSMFARNIYQLSFRQN--ASDKEIVWVMRITVFVFGASATAM 395
      || ||:||||| :| || | :||| ||:| | :::| : | : :
Db    329 VAVLAAIMSSADSQLLVGSSAFVRDIYQKMERRNRKLSQKKLVRLSRLTTVVFMGLSLIL 388

Qy    396 ALLTKTVYGLWYLSSDLVYIVIF-----PQLLCVLVFKGTNTYGAVAGYVSGLFL 445
      | | : | :|| | || : || | : | :||
Db    389 A-FTAQEFVFW-----MVLFAFGGLGACFGPALLLSFYWKGVTROGVWLGMIAGLLT 439

Qy    446 RITGGEPYLYLQPLIFYPGYYPDDNGIYNQKFPFKTLAMVTSFLT----NICISYLAK 499
      | : || | : || : | | | | : : || |
Db    440 VI-----LVKQQPQWTY-AFLPDVKELLNTYFFGITYEAVPGFIVATTITVVISLFTK 491

```

RESULT 10

A42251

nucleoside transport protein - rabbit

N;Alternate names: Na<sup>+</sup>/nucleoside cotransporter, SNST1

C;Species: Oryctolagus cuniculus (domestic rabbit)

C;Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 20-Aug-1999

C;Accession: A42251

R;Pajor, A.M.; Wright, E.M.  
 J. Biol. Chem. 267, 3557-3560, 1992  
 A;Title: Cloning and functional expression of a mammalian Na<sup>+</sup>/nucleoside cotransporter. A member of the SGLT family.  
 A;Reference number: A42251; MUID:92156077; PMID:1740408  
 A;Accession: A42251  
 A;Molecule type: mRNA  
 A;Residues: 1-672 <PAJ>  
 A;Cross-references: GB:M84020; NID:g165550; PIDN:AAA31421.1; PID:g165551  
 A;Note: sequence extracted from NCBI backbone (NCBIN:82253, NCBIP:82256)  
 C;Superfamily: proline carrier protein  
 C;Keywords: membrane protein; nucleoside transport

Query Match 10.0%; Score 298; DB 2; Length 672;  
 Best Local Similarity 25.0%; Pred. No. 4e-14;  
 Matches 153; Conservative 89; Mismatches 232; Indels 138; Gaps 25;

```

Qy      9  IAIIVFYLLILLVGIWAAWRTKNSGSAERSEAIIVGGRDIGLLVGGETMTATWVGGGY 67
      ||:|  ::||:: ||:|  || |  | :      : || :  | :: |: :| | :
Db     26  IAVIAAYFLLVIGVGLWSMCRT-NRGTV----GGYFLAGRSMVWVPVGASLFASNIGSGH 80

Qy     68  INGTAEEVYVPGYGLAWAQAPIGYSLS----LILGGLFFAKPMRSKGYVTMLDPFQQIYG 123
      ||  ||  ||  |  | :  :  ::| |  ||  :  | :||  |
Db     81  FVGLA-----GTGAANGLAVAGFEWNALFVVLGLWLFAPVYLTAGVITM----PQYLR 130

Qy    124  KRMGG-----LLFIPALMGEMFWAAAI--SALGATISVIIDVDMHISVIISA 169
      || ||  ||:|  :  :: |:|  || |  |  : |||
Db    131  KRFGGHRIRLYLSVLSFLYIFTKISVDMFSGAVFIQQALGWN-----YASVIAL 182

Qy    170  LIATLYTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKY----- 224
      |  :||: ||| :: ||| || | | | :|  :|  |:: : ||
Db    183  GITMVYTVTGGLAALMYTDTVQTFVIIAGAFILTGYAFHEVG----GYSGLFDKYMGAMT 238

Qy    225  -----QKPWLGTVDSSSEVSWLDSFLLL---MLGGIPW-----QAYF 258
      : | :| : ||  ||: ||  : | :||  |
Db    239  SLTVSEDPVAGNISSSCYRPRPDSYHLLRDPVTGDLWPALLGLTIVSGWYWCSDQVIV 298

Qy    259  QRVLSSSSATYAQVLSFLAAFGLVMAIPAILIGAIGASTDWNQTAYGLPDKT----TE 314
      || |:  : |: :  | : |  :: | |  :: |  |:  ||
Db    299  QRCLAGRNLTTHIKAGCILCGYLKLTMPFLMVMPGMISRILYPDEVACVAPEVCKRVCSTE 358

Qy    315  E--ADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFARNIYQLSFRQNA 372
      ::|  | :: | | :  | : ||: || |  | :|::|  :|| | | |
Db    359  VGCSNIAYPRLVVKLMPNGLRGLMLAVMLAALMSSLASIFNSSSTLFTMDIYTL--RPRA 416

Qy    373  SDKEIVWVMRITVFVFGASATAMALLTKTVYG---LWYLSSDLVYIV--IFPQLLCVLFV 427
      : |:| | |: |  | : |  : :  |  |: |  | :  :  |||
Db    417  GEGELLVGRWVVFIVAVSVAWLPVVQAAQGGQLFDYIQSVSSYLAPPVSAVFVVALFV 476

Qy    428  KGTNTYGAVAGYVSGFLFRITGGEPLYLYLQPLIFYPGYYPDDNGIYNQKFPFKTLAMV-- 485
      | ||  | : || : :  |  : || | : |
Db    477  PRVNEKGAFWGLIGLLMGLARLIP-----EFSFGTGSCVRP 513

Qy    486  ---TSFLTNICISYLAKYLFE-SG-----TLP-PKLDVFDVAVVA-RHSEENMDKTI 530
      :|| :  | | || ||  | | | :  | : ||:|
Db    514  SACPAFLCRVHYLYFAIVLFFCSGLLIIIVSLCTAPIPRKHLHRLVFSRLHSKE----- 567

```

Qy 531 LVKNENIKLDEL 542  
: |:: |||  
Db 568 --EREDLDADEL 577

RESULT 11

S59637

glucose transport protein SGLT1, intestinal - sheep

N;Alternate names: Na+/glucose cotransporter SGLT1

C;Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)

C;Date: 10-Apr-1996 #sequence\_revision 19-Apr-1996 #text\_change 20-Aug-1999

C;Accession: S59637; S48858

R;Tarpey, P.S.; Wood, I.S.; Shirazi-Beechey, S.P.; Beechey, R.B.

Biochem. J. 312, 293-300, 1995

A;Title: Amino acid sequence and the cellular location of the Na(+)-dependent D-glucose symporters (SGLT1) in the ovine enterocyte and the parotid acinar cell.

A;Reference number: S59637; MUID:96077158; PMID:7492327

A;Accession: S59637

A;Status: nucleic acid sequence not shown

A;Molecule type: mRNA

A;Residues: 1-664 <TAR>

A;Cross-references: EMBL:X82411; NID:g861072; PIDN:CAA57809.1; PID:g861073

A;Experimental source: tissue type jejunal mucosa,

R;Wood, I.

submitted to the EMBL Data Library, October 1994

A;Reference number: S48858

A;Accession: S48858

A;Molecule type: mRNA

A;Residues: 1-233, 'R', 235-432, 'V', 434-466, 'MR', 469-664 <WOO>

A;Cross-references: EMBL:X82411

C;Superfamily: proline carrier protein

Query Match 9.9%; Score 294; DB 2; Length 664;  
Best Local Similarity 23.9%; Pred. No. 7.7e-14;  
Matches 127; Conservative 93; Mismatches 202; Indels 110; Gaps 23;

Qy 11 IIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGY--- 67  
|:~::~: ||:| | : | | | : : || : | : : | | :  
Db 32 IVIYFVVVMAVGLWAMFST-NRGTV----GGFFLAGRSMVWWPIGASLFASNIGSGHFVG 86  
Qy 68 INGTAEEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSK-GYVTMLDPFQQIYGKRM 126  
: || | : | | | : || : | | | : ||  
Db 87 LAGTGAAAGIATGGFEWN----ALILVVLWGWFV--PIYIKAGVVTM----PEYLRKRF 136  
Qy 127 GG-----LLFIPALMGEMFWAAAI FSALGATISVIIDVDMHISVIISALIATL 174  
|| : | : : || | | : : : | | |  
Db 137 GGQRIQVYLSVLSLVLYIFTKISADIFSGAIF-----INLALGLDLYLAIFILLAITAL 190  
Qy 175 YTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDS 234  
||: ||| : | ||| : | : : | || | : | | : |||  
Db 191 YTITGGLAAVIYTDTLQTVMILGFSILTGFASFHEVG----GYSAFVTKYMNA-IPTVTS 245  
Qy 235 -----SEVYS-WLDSFLLL---MLGGIPW-----QAYFQRLSSS 265  
| | : ||| : : | : || | || ||:  
Db 246 YGNTTVKKECYTPRADS FHIRDPLKGDLPWPGLIFGLTIISLWYWCTDQVIVQRCLSAK 305  
Qy 266 SATYAQVLSFLAAGCLVMAIPAILIGAIGASTDWNQTAYGLPDPKTTEE-----AD 317

```

      : :: :   : : | :   : : | |   : | : | | :   :
Db      306 NMSHVKAGCIMCGYMKLLPMFLMVMPGMISRILFTEKVACTV--PSECEKYCGTKVGCTN 363

Qy      318 MILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFARNIYQLSFRQNASDKEI 377
      : | :: | | :   | : | :: || | | || :: | : | | : | :: || :
Db      364 IAYPTLVVEIMPNGLRGLMLSVMLASLMSSITSIFNSASTLFTMDIY-TKIRKKASEKEL 422

Qy      378 VWVMRITVFV-FGASATAMALLTKTVYG--LWYLSSDLVYI--VIFPQLLCVLFVKGTNT 432
      : | : : | | | : :: | | : | | : | | : | | |
Db      423 MIAGRLEMLVLIGVSIWVPIVQSAQSGQLFDYIQSITSYLGPPIAAVFLLAIFCKRVNE 482

Qy      433 YGAVAGYVSGFLFLRI-----TG-----GEPYLYLQPLIF 461
      || | : | : :   ||   | || | :: |
Db      483 PGAFWGLIIGFLIGVSRMITEFAYGTGSCMEPSNCPTIICGVHYLYFAILF 534

```

# RESULT 12

A56765

sodium-glucose cotransporter homolog - human

C;Species: Homo sapiens (man)

C;Date: 08-Sep-1995 #sequence\_revision 08-Sep-1995 #text\_change 20-Aug-1999

C;Accession: A56765; I51890

R;Wells, R.G.; Pajor, A.M.; Kanai, Y.; Turk, E.; Wright, E.M.; Hediger, M.A.

Am. J. Physiol. 263, F459-F465, 1992

A;Title: Cloning of a human kidney cDNA with similarity to the sodium-glucose cotransporter.

A;Reference number: A56765; MUID:93035768; PMID:1415574

A;Accession: A56765

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-672 <WEL>

A;Cross-references: GB:M95549; NID:g338052; PIDN:AAA36608.1; PID:g338053

A;Experimental source: kidney cortex

C;Superfamily: proline carrier protein

C;Keywords: transmembrane protein

```

Query Match          9.8%; Score 292; DB 2; Length 672;
Best Local Similarity 24.1%; Pred. No. 1.1e-13;
Matches 147; Conservative 91; Mismatches 237; Indels 136; Gaps 22;

```

```

Qy      8 LIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGY 67
      :: | : || :: || : | | | :   : || : | :: | : | | :
Db      26 ILVIAAYFLLVIGVGLWSMCRT-NRGTV----GGYFLAGRSMVWVPVGASLFASNIGSGH 80

Qy      68 INGTAEAVYVPGYGLAWAQAPIGYLSL----LILGGLFFAKPMRSKGYVTMLDPFQQIYG 123
      | |   | | | | | : :   : : | | | : | : | |   |
Db      81 FVGLA-----GTGAASGLAVAGFEWNALFVLLLGWLFAPVYLTAGVITM----PQYLR 130

Qy      124 KRMGG-----LLFIPALMGEMFWAAAI--SALGATISVIIDVDMHISVIISA 169
      || ||   | : | :   : : | : | || | |   : || |
Db      131 KRFGGRRIRLYLSVLSLFLYIFTKISVDMFSGAVFIQQALGWN-----YASVIAL 182

Qy      170 LIATLYTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKY----- 224
      | : || : || | : || | | | | | : : | | : : | |
Db      183 GITMIYTVTGGLAALMYTDTVQTFVILGGACILMGYAFHEVG----GYSGLFDKYLGAAT 238

Qy      225 -----QKPWLGTVDSSSEVYSWLDSELLL---MLGGIPW-----QAYF 258

```





```

Db      32 IVIYFVVVMAVGLWHMFST-NRGTV----GGFFLAGRSMVWWPIGASLFASNIGSGHFVG 86
Qy      68 INGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSK-GYVTMLDPFQQIYGKRM 126
      : || | : | | | : || : | : | | | : ||
Db      87 LAGTGAAAGIATGGFEWN----ALILVLLGWVFV--PIYIKAGVVTM----PEYLRKRF 136
Qy     127 GG-----LLFIPALMGEMFWAAAFSALGATISVIIDVDMHISVIISALIATL 174
      || : || : : || | : : : | : : : | | |
Db     137 GGQRIQVYLSVLSLVLYIFTKISADIFSGAIF-----INLALGLDLYLAI FILLAITAL 190
Qy     175 YTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDS 234
      || : || | : | || : | : : | | | : || : || |
Db     191 YTITGGLAAVIYTDTLQTVIMLLGSFILTGFAFHEVG----GYSAFVTKYMNA-IPTVTS 245
Qy     235 -----SEVYS-WLDSFLLL---MLGGIPW-----QAYFQRLSSS 265
      | | : || | : : | : || | | | | :
Db     246 YGNTTVKKECYTPRADSFHIFRDPLKGDLPWPGLIFGLTIISLWYWCTDQVIVQRCLSAK 305
Qy     266 SATYAQVLSFLAAGCLVMAIPAILIGAIGASTDWNQTAYGLPDPKTTEE-----AD 317
      : : : : : | : : : | | : | : | | :
Db     306 NMSHVKAGCIMCGYMKLLPMFLMMPGMISRILFTEKVACTV--PSECEKYCGTKVGCTN 363
Qy     318 MILPIVLQYLCPVYISFFGLGAVSAAMVSSADSSILSASSMFARNIYQLSFRQNASDKEI 377
      : | : : | | : | : | : || | | || : | : | : | :
Db     364 IAYPTLVVELMPNGLRGLMLSVMLASLMSLTSIFNSASTLFTMDIY-TKIRKKASEKEL 422
Qy     378 VWVMRITVFV-FGASATAMALLTKTVYG--LWYLSSDLVYI--VIFPQLLCVLFVKGTNT 432
      : | : : | | | : : : | | : | : | | | |
Db     423 MIAGRLFMLVLIGVSIWVPIVQSAQSGQLFDYIQSITSYLGPPPIAAVFLLAIFCKRVNE 482
Qy     433 YGAVAGYVSGFLFLRI-----TG-----GEPYLYLQPLIF 461
      || | : | : : || | || | : |
Db     483 PGAFWGLIIGFLIGVSRMITEFAYGTGSCMEPSNCPTIICGVHYLYFAIILF 534

```

# RESULT 14

H71097

hypothetical protein PH1044 - *Pyrococcus horikoshii*

C;Species: *Pyrococcus horikoshii*

C;Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 20-Jun-2000

C;Accession: H71097

R;Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kosugi, H.; Hosoyama, A.; Nagai, Y.; Sakai, M.; Ogura, K.; Otsuka, R.; Nakazawa, H.; Takamiya, M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.; Horikoshi, K.; Masuchi, Y.; Shizuya, H.; Kikuchi, H. DNA Res. 5, 55-76, 1998

A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, *Pyrococcus horikoshii* OT3.

A;Reference number: A71000; MUID:98344137; PMID:9679194

A;Accession: H71097

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-491 <KAW>

A;Cross-references: GB:AP000004; NID:g3236131; PIDN:BAA30142.1; PID:g3257459

A;Experimental source: strain OT3

A;Note: this accession replaces an interim accession for a sequence replaced by GenBank  
 C;Genetics:  
 A;Gene: PH1044  
 C;Superfamily: proline carrier protein

Query Match 9.6%; Score 286; DB 2; Length 491;  
 Best Local Similarity 22.9%; Pred. No. 2.1e-13;  
 Matches 125; Conservative 83; Mismatches 197; Indels 142; Gaps 20;

```

Qy      8 LIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGY 67
      || :: : | |::| : | : | : | | :||: |
Db     17 LIIVLGWVFLSLIVGMAGIKRKFT-----LEGYLVSGRTLGLIFLYVLMAGEIYSAYA 70

Qy     68 INGTAEEVYVPGYGLAWAQAPIGYSLSLILGGLFFA---KPMRSKGYVTMLDPFQQIYG 123
      || | | : :| || | | :| | :| || | | |
Db     71 FLGTGGWAYSYGMPIMYA---IGYGALAYSFGYFYARYVWKAGKAFGCVTQADYFQVRYN 127

Qy    124 KRMGGLLFIPALMGEMF---WAAAI FSALGATISV--IIDVDMHISVIISALIATLYTLV 178
      : | : ||:| : | : | : | : : :| :| :|
Db    128 SK--ALAVLVALIGIIFNIPYLQLQLQGLGYIVHVGSLGSITPKAGIVIGMIIMMIYVYT 185

Qy    179 GGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSEVY 238
      || :::|:::| :|: | :| | : | | ||
Db    186 SGLRGISWTNLLQATLMFIVAWV-VLFTIPFKQFGGIGELFKTLAQTTP----- 233

Qy    239 SWLDSFLLLMLGGIPWQAYFQRLSSSSATYAQVLSFLAAGCLV--MAIPAILIGAIGA 296
      | :| || | | | | : | : | :|
Db    234 ---DHLILHPPLGISW-----YVSTL-ILSGLGFFMYPQLYPSI----- 268

Qy    297 STDWNQTAYGLPDPKTTEEADMILPIVLQYLCPVYISFFGLGAVS----- 341
      || | || : :||: : || :| :| :|
Db    269 -----YGARDLKTLLKRNIVLLPLYSIFMIPVILAGFTVAALGIKLSAPDEAVLKAVE 320

Qy    342 -----AAVMSSADSSILSASSMFARNIYQLSFRQNASDKEIVVWMRITV 385
      || | :| : :|| : : :|:| : : | |||:| | ||:|
Db    321 ITYPSWVLGVGAAGFAAAASTASAILLSLAGLLSKNLYAIA-KPTASDKELVLVSRISV 379

Qy    386 FVFGASATAMALLTKTVYGLWYLSDDLVIYIVI-----FPQLLCVLFVKGTNTY--- 433
      : | | :|| | || ::| | : || | | |
Db    380 ILLGLLAMGLAL-----YSPGRLVSLLLAYAGMTQMFP GAVFGLFWKRMNKYATG 430

Qy    434 -GAVAGYVSGLFLRITGGEPYLYLQPLIFYPGYYPDDNGIYNQKFPFKTLAMVTSFLTNI 492
      | :|| : :||: | || ||: | : : : :
Db    431 TGIIAGLITVAYLRLV----LKKNP-----GIH----FGLWGLLVNIIVTL 469

Qy    493 CISYLAK 499
      ::|| |
Db    470 IVAYLTK 476
  
```

# RESULT 15

H69670

sodium/proline symporter opuE - *Bacillus subtilis*

N;Alternate names: proline transporter opuE

C;Species: *Bacillus subtilis*

C;Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 20-Jun-2000

C;Accession: H69670; T44450

R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.; Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans, A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.; Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.

Nature 390, 249-256, 1997

A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim, S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.; Guiseppi, G.; Guy, B.J.; Haga, K.; Haiech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.; Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.

A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Porwollik, S.; Prescott, A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.

A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.; Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenegger, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.

A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.

A;Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.

A;Reference number: A69580; MUID:98044033; PMID:9384377

A;Accession: H69670

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-492 <KUN>

A;Cross-references: GB:Z99107; GB:AL009126; NID:g2632866; PIDN:CAB12486.1; PID:g2632980

A;Experimental source: strain 168

R;Borriss, R.

submitted to the EMBL Data Library, June 1997

A;Reference number: Z22776

A;Accession: T44450

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-492 <BOR>

A;Cross-references: EMBL:AF011545; PIDN:AAB72182.1

C;Genetics:

A;Gene: opuE

A;Map position: 56 degree

C;Function:

A;Description: catalyzes the uptake of proline by a Na<sup>+</sup>-dependent transport mechanism

C;Superfamily: proline carrier protein

```
F;44-71/Domain: transmembrane #status predicted <TM1>
F;126-145/Domain: transmembrane #status predicted <TM2>
F;161-183/Domain: transmembrane #status predicted <TM3>
F;189-208/Domain: transmembrane #status predicted <TM4>
F;231-253/Domain: transmembrane #status predicted <TM5>
F;272-295/Domain: transmembrane #status predicted <TM6>
F;311-347/Domain: transmembrane #status predicted <TM7>
F;366-386/Domain: transmembrane #status predicted <TM8>
F;391-417/Domain: transmembrane #status predicted <TM9>
F;422-438/Domain: transmembrane #status predicted <TM10>
F;452-470/Domain: transmembrane #status predicted <TM11>
```

Qy	5	VEGLIAIVFYLLILLVGIWAARWTKNSGSAERSEAIIVGGRDGLLLVGGFTMTATWVG	64
Db	3	IEIIISLGIYFIAMLLIGWYAFKKTDDIND-----YMLGGRGLGPFVTALSAGAADMS	55
Qy	65	GGYINGTAEAVYVPGYGLAWAQAPI---GYSLSLILGGLFFAKPMRSKGYVTMLDPFQOI	121
Db	56	GWMLMGVPGAMFATGLSTLWLALGLTIGAYSNYLLAPRLRAYTEAADDAITIPDFFDKR	115
Qy	122	YGKRMGGLLFIPALMGEMFWAAAIIFSAL---GATISVIIDVDMHISVIISALIATLYTLV	178
Db	116	FQHSSSLKIVSALIIMIFFTLYTSSGMVSGGRLFESAFGADYKLGLELTAVVLYTLF	175
Qy	179	GGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSEVY	238
Db	176	GGFLAVSLTDFVQGAIMFAAL-VLVPI-----VAFT--HVGGVAPTfHEIDAVNPH	223
Qy	239	SWLDSF----LLLMLGGIPWQAYFQRVLSSSSATYAQ---VLSFLA-----	277
Db	224	-LLDIFKGASVISIISYLAWGLGY-----YGQPHIIVRFMAIKDKLKPARRIG	272
Qy	278	AFGCLVMAIPAILIGAIGASTDWNQTAYGLPDPKTTEEADMILPIVLQYLCVPYISFFGL	337
Db	273	MSWMIITVLGSVLTGLIG-----VAYAHKFGVAVKDPEMIFIIFSKILFHPLITGFL	325
Qy	338	GAVSAAVMSSADSSILSASSMFARNIYQLSFRQNASDKEIVWVMRITVVFVGASATAMAL	397
Db	326	SAILAAIMSSISSQLLVITASAVTEDLYRSFFRRKASDKELVMIGRLSVLVIHAVIALLSL	385
Qy	398	LTKTVYGLWYLLSSDLVYIVIF-----PQLLCVLFVKGTNTYGAVAGYVSG---LF	444
Db	386	NP-----NSTILDVGAYAWAGFGSAFGPAILLSLYWKRMNEWGALAAMIVGAATVL	436
Qy	445	LRITGGEPYLYLQPLIFYPGYYPDDNGIYNQKFPFKTLAMVTSFLTNICISYLA	499
Db	437	IWITTG-----IAKSTGVY-EIIP----GFILSMIAGIIVSMITK	471

Search completed: March 22, 2004, 15:35:51  
Job time : 44 secs

OM protein - protein search, using sw model

Run on: March 22, 2004, 15:32:29 ; Search time 531 Seconds  
(without alignments)  
282.851 Million cell updates/sec

Title: US-10-069-541-6  
Perfect score: 2972  
Sequence: 1 MAFHVEGLIAIIVFYLLILL.....EAFLDVDSSPEGSGTEDNLQ 580

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1049977 seqs, 258955339 residues

Total number of hits satisfying chosen parameters: 1049977

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep:\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep:\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep:\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep:\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result	Score	Match	Length	DB	ID	Description
No.						

1	2972	100.0	580	10	US-09-911-077A-2	Sequence 2, Appli
2	2972	100.0	580	10	US-09-911-077A-10	Sequence 10, Appl
3	2972	100.0	580	10	US-09-911-077A-11	Sequence 11, Appl
4	2972	100.0	580	10	US-09-911-077A-12	Sequence 12, Appl
5	2820	94.9	580	10	US-09-911-077A-6	Sequence 6, Appli
6	2795	94.0	580	10	US-09-911-077A-4	Sequence 4, Appli
7	2795	94.0	580	10	US-09-911-077A-24	Sequence 24, Appl
8	1506.5	50.7	610	12	US-10-241-784-2	Sequence 2, Appli
9	1453	48.9	576	10	US-09-911-077A-8	Sequence 8, Appli
10	311.5	10.5	675	9	US-09-733-630-2	Sequence 2, Appli
11	306	10.3	486	14	US-10-156-761-12818	Sequence 12818, A
12	306	10.3	664	14	US-10-119-988-12	Sequence 12, Appl
13	298.5	10.0	675	9	US-09-928-530-2	Sequence 2, Appli
14	298.5	10.0	675	14	US-10-162-012-27	Sequence 27, Appl
15	298.5	10.0	675	15	US-10-162-102-27	Sequence 27, Appl
16	297.5	10.0	471	12	US-10-282-122A-52725	Sequence 52725, A
17	295	9.9	596	14	US-10-119-988-8	Sequence 8, Appli
18	292	9.8	672	9	US-09-928-530-5	Sequence 5, Appli
19	292	9.8	672	14	US-10-162-012-30	Sequence 30, Appl
20	292	9.8	672	15	US-10-162-102-30	Sequence 30, Appl
21	287	9.7	678	12	US-10-072-012-438	Sequence 438, App
22	281.5	9.5	673	12	US-10-072-012-440	Sequence 440, App
23	279	9.4	454	12	US-10-282-122A-53545	Sequence 53545, A
24	277.5	9.3	596	9	US-09-740-026A-2	Sequence 2, Appli
25	277.5	9.3	596	12	US-10-072-012-114	Sequence 114, App
26	277.5	9.3	596	12	US-10-169-395-124	Sequence 124, App
27	277.5	9.3	596	14	US-10-237-859-2	Sequence 2, Appli
28	277.5	9.3	643	14	US-10-119-988-5	Sequence 5, Appli
29	277	9.3	596	9	US-09-740-026A-4	Sequence 4, Appli
30	277	9.3	596	14	US-10-237-859-4	Sequence 4, Appli
31	277	9.3	597	12	US-10-072-012-436	Sequence 436, App
32	272.5	9.2	524	9	US-09-738-626-6949	Sequence 6949, Ap
33	272.5	9.2	524	12	US-10-627-476-496	Sequence 496, App
34	269.5	9.1	612	12	US-10-072-012-116	Sequence 116, App
35	269.5	9.1	664	14	US-10-119-988-2	Sequence 2, Appli
36	262.5	8.8	477	12	US-10-282-122A-67179	Sequence 67179, A
37	262	8.8	512	15	US-10-161-493-36	Sequence 36, Appl
38	260	8.7	552	12	US-10-072-012-439	Sequence 439, App
39	260	8.7	674	14	US-10-173-123-9	Sequence 9, Appli
40	260	8.7	681	14	US-10-173-123-7	Sequence 7, Appli
41	260	8.7	752	15	US-10-297-022-20	Sequence 20, Appl
42	254	8.5	561	12	US-10-072-012-118	Sequence 118, App
43	249	8.4	479	12	US-10-282-122A-46830	Sequence 46830, A
44	248	8.3	742	15	US-10-297-022-15	Sequence 15, Appl
45	245	8.2	738	14	US-10-173-123-13	Sequence 13, Appl

#### ALIGNMENTS

RESULT 1  
 US-09-911-077A-2  
 ; Sequence 2, Application US/09911077A  
 ; Publication No. US20030114399A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: BLAKELY, RANDY D.

```
; APPLICANT: APPARSUNDARAM, SUBRAMANIAM
; APPLICANT: FERGUSON, SHAWN
; TITLE OF INVENTION: HUMAN AND MOUSE CHOLINE TRANSPORTER cDNA
; FILE REFERENCE: VBLT:008US
; CURRENT APPLICATION NUMBER: US/09/911,077A
; CURRENT FILING DATE: 2001-07-23
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 580
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-911-077A-2
```

```
Query Match          100.0%; Score 2972; DB 10; Length 580;
Best Local Similarity 100.0%; Pred. No. 2.6e-269;
Matches 580; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1 MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTA 60
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTA 60

Qy     61 TWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ 120
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ 120

Qy    121 IYGKRMGGLLFIPALMGEMFWAAAI FSALGATISVIIDVDMHISVIISALIATLYTLVGG 180
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 IYGKRMGGLLFIPALMGEMFWAAAI FSALGATISVIIDVDMHISVIISALIATLYTLVGG 180

Qy    181 LYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSSEVYSW 240
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 LYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSSEVYSW 240

Qy    241 LDSFLLMLGGIPWQAYFQRLSSSSATYAQVLSFLAAGCLVMAIPAILIGAIGASTDW 300
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 LDSFLLMLGGIPWQAYFQRLSSSSATYAQVLSFLAAGCLVMAIPAILIGAIGASTDW 300

Qy    301 NQTAYGLPDPKTTTEADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFA 360
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 NQTAYGLPDPKTTTEADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFA 360

Qy    361 RNIYQLSFRQNASDKEIVVMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQ 420
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 RNIYQLSFRQNASDKEIVVMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQ 420

Qy    421 LLCVLFVKGTNTYGA VAGYVSGFLRITGGEPYLYLQPLIFYPGYPPDDNGIYNQKFPPK 480
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 LLCVLFVKGTNTYGA VAGYVSGFLRITGGEPYLYLQPLIFYPGYPPDDNGIYNQKFPPK 480

Qy    481 TLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDVAVVARHSEENMDKTILVKENIKLD 540
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDVAVVARHSEENMDKTILVKENIKLD 540

Qy    541 ELALVKPRQSMTLSSTFTNKEAFLDVDSSPEGSGTEDNLQ 580
      ||||||||||||||||||||||||||||||||||||||||||
Db    541 ELALVKPRQSMTLSSTFTNKEAFLDVDSSPEGSGTEDNLQ 580
```



RESULT 2

US-09-911-077A-10

; Sequence 10, Application US/09911077A  
; Publication No. US20030114399A1  
; GENERAL INFORMATION:  
; APPLICANT: BLAKELY, RANDY D.  
; APPLICANT: APPARSUNDARAM, SUBRAMANIAM  
; APPLICANT: FERGUSON, SHAWN  
; TITLE OF INVENTION: HUMAN AND MOUSE CHOLINE TRANSPORTER cDNA  
; FILE REFERENCE: VBLT:008US  
; CURRENT APPLICATION NUMBER: US/09/911,077A  
; CURRENT FILING DATE: 2001-07-23  
; NUMBER OF SEQ ID NOS: 27  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 10  
; LENGTH: 580  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-911-077A-10

Query Match 100.0%; Score 2972; DB 10; Length 580;  
Best Local Similarity 100.0%; Pred. No. 2.6e-269;  
Matches 580; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTA	60
Db	1	MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTA	60
Qy	61	TWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ	120
Db	61	TWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ	120
Qy	121	IYGKRMGGLLFIPALMGEMFWAAAFSALGATISVIIDVDMHISVIISALIATLYTLVGG	180
Db	121	IYGKRMGGLLFIPALMGEMFWAAAFSALGATISVIIDVDMHISVIISALIATLYTLVGG	180
Qy	181	LYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSSEVYSW	240
Db	181	LYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSSEVYSW	240
Qy	241	LDSFLLMLGGIPWQAYFQRVLSSSSATYAQVLSFLAAGCLVMAIPAILIGAIGASTDW	300
Db	241	LDSFLLMLGGIPWQAYFQRVLSSSSATYAQVLSFLAAGCLVMAIPAILIGAIGASTDW	300
Qy	301	NQTAYGLPDPKTTEEADMILPIVLQYLCVPYISFFGLGAVSAAVMSSADSSILSASSMFA	360
Db	301	NQTAYGLPDPKTTEEADMILPIVLQYLCVPYISFFGLGAVSAAVMSSADSSILSASSMFA	360
Qy	361	RNIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQ	420
Db	361	RNIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQ	420
Qy	421	LLCVLFVKGTNTYGA VAGYVSGFLFRITGGEPYLYLQPLIFYPGYPPDDNGIYNQKFPFK	480
Db	421	LLCVLFVKGTNTYGA VAGYVSGFLFRITGGEPYLYLQPLIFYPGYPPDDNGIYNQKFPFK	480

```

Qy      481 TLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDVAVVARHSEENMDKTILVKNENIKLD 540
         |||
Db      481 TLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDVAVVARHSEENMDKTILVKNENIKLD 540

Qy      541 ELALVKPRQSMTLSSTFTNKEAFLDVDSSPEGSGTEDNLQ 580
         |||
Db      541 ELALVKPRQSMTLSSTFTNKEAFLDVDSSPEGSGTEDNLQ 580

```

RESULT 3

US-09-911-077A-11

```

; Sequence 11, Application US/09911077A
; Publication No. US20030114399A1
; GENERAL INFORMATION:
; APPLICANT: BLAKELY, RANDY D.
; APPLICANT: APPARSUNDARAM, SUBRAMANIAM
; APPLICANT: FERGUSON, SHAWN
; TITLE OF INVENTION: HUMAN AND MOUSE CHOLINE TRANSPORTER cDNA
; FILE REFERENCE: VBLT:008US
; CURRENT APPLICATION NUMBER: US/09/911,077A
; CURRENT FILING DATE: 2001-07-23
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 580
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-911-077A-11

```

```

Query Match          100.0%; Score 2972; DB 10; Length 580;
Best Local Similarity 100.0%; Pred. No. 2.6e-269;
Matches 580; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1 MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTA 60
         |||
Db      1 MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTA 60

Qy      61 TWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ 120
         |||
Db      61 TWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ 120

Qy      121 IYGKRMGGLLFIPALMGEMFWAAAI FSALGATISVIIDVDMHISVIISALIATLYTLVGG 180
         |||
Db      121 IYGKRMGGLLFIPALMGEMFWAAAI FSALGATISVIIDVDMHISVIISALIATLYTLVGG 180

Qy      181 LYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSEVYSW 240
         |||
Db      181 LYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSEVYSW 240

Qy      241 LDSFLLMLGGIPWQAYFQRLSSSSATYAQVLSFLAAGCLVMAIPAILIGAIGASTDW 300
         |||
Db      241 LDSFLLMLGGIPWQAYFQRLSSSSATYAQVLSFLAAGCLVMAIPAILIGAIGASTDW 300

Qy      301 NQTAYGLPDPKTTTEADMILPIVLQYLCVPYISFFGLGAVSAAVMSSADSSILSASSMFA 360
         |||
Db      301 NQTAYGLPDPKTTTEADMILPIVLQYLCVPYISFFGLGAVSAAVMSSADSSILSASSMFA 360

```

```

Qy      361 RNIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQ 420
          |||
Db      361 RNIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQ 420

Qy      421 LLCVLFVKGTNTYGAVAGYVSGFLRITGGEPYLYLQPLIFYPGYPPDDNGIYNQKFPFK 480
          |||
Db      421 LLCVLFVKGTNTYGAVAGYVSGFLRITGGEPYLYLQPLIFYPGYPPDDNGIYNQKFPFK 480

Qy      481 TLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDVAVVARHSEENMDKTILVKNENIKLD 540
          |||
Db      481 TLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDVAVVARHSEENMDKTILVKNENIKLD 540

Qy      541 ELALVKPRQSM T L S S T F T N K E A F L D V D S S P E G S G T E D N L Q 580
          |||
Db      541 ELALVKPRQSM T L S S T F T N K E A F L D V D S S P E G S G T E D N L Q 580

```

RESULT 4

US-09-911-077A-12

; Sequence 12, Application US/09911077A

; Publication No. US20030114399A1

; GENERAL INFORMATION:

; APPLICANT: BLAKELY, RANDY D.

; APPLICANT: APPARSUNDARAM, SUBRAMANIAM

; APPLICANT: FERGUSON, SHAWN

; TITLE OF INVENTION: HUMAN AND MOUSE CHOLINE TRANSPORTER cDNA

; FILE REFERENCE: VBLT:008US

; CURRENT APPLICATION NUMBER: US/09/911,077A

; CURRENT FILING DATE: 2001-07-23

; NUMBER OF SEQ ID NOS: 27

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 12

; LENGTH: 580

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-911-077A-12

```

Query Match          100.0%;  Score 2972;  DB 10;  Length 580;
Best Local Similarity 100.0%;  Pred. No. 2.6e-269;
Matches 580;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

```

```

Qy      1 MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTA 60
          |||
Db      1 MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTA 60

Qy      61 TWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ 120
          |||
Db      61 TWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ 120

Qy      121 IYGKRMGGLLFIPALMGEMFWAAAFSALGATISVIIDVDMHISVIISALIATLYTLVGG 180
          |||
Db      121 IYGKRMGGLLFIPALMGEMFWAAAFSALGATISVIIDVDMHISVIISALIATLYTLVGG 180

Qy      181 LYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSSEVYSW 240
          |||
Db      181 LYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSSEVYSW 240

```

Qy	241	LDSFLLMLGGIPWQAYFQRVLSSSSATYAQVLSFLAAFGLVMAIPAILIGAIGASTDW	300
Db	241	LDSFLLMLGGIPWQAYFQRVLSSSSATYAQVLSFLAAFGLVMAIPAILIGAIGASTDW	300
Qy	301	NQTAYGLPDPKTTTEADMILPIVLQYLCVPYISFFGLGAVSAAVMSSADSSILSASSMFA	360
Db	301	NQTAYGLPDPKTTTEADMILPIVLQYLCVPYISFFGLGAVSAAVMSSADSSILSASSMFA	360
Qy	361	RNIYQLSFRQNASDKEIVWMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQ	420
Db	361	RNIYQLSFRQNASDKEIVWMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQ	420
Qy	421	LLCVLFVKGTNTYGAVAGYVSGFLRITGGEPYLYLQPLIFYPGYYPDDNGIYNQKFPFK	480
Db	421	LLCVLFVKGTNTYGAVAGYVSGFLRITGGEPYLYLQPLIFYPGYYPDDNGIYNQKFPFK	480
Qy	481	TLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDVAVVARHSEENMDKTILVKENIKLD	540
Db	481	TLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDVAVVARHSEENMDKTILVKENIKLD	540
Qy	541	ELALVKPRQSMTLSSTFTNKEAFLDVDSSPEGSGTEDNLQ	580
Db	541	ELALVKPRQSMTLSSTFTNKEAFLDVDSSPEGSGTEDNLQ	580

```
Query Match          94.9%;   Score 2820;   DB 10;    Length 580;  
Best Local Similarity 93.1%;   Pred. No. 4.5e-255;  
Matches  540; Conservative  24; Mismatches  16; Indels      0; Gaps       0
```

  

```
QY           1 MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFMTTA 60  
              | |||||:::||:|||| |||||::|||:|||||||  
Db            1 MPFHVLEGLVAIIIFYLLIFLVGIWAAWKTKNSGNAEERSEAIIVGGRDIGLLVGGFMTTA 60
```

  

```
QY           61 TWVGGGYINGTAEAVYPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKG YVTMLDPFQQ 120  
              ||||| || || |||||  
Db            61 TWVGGGYINGTAEAVYGPGCGLAWAQAPIGYSLSLILGGLFFAKPMRSKG YVTMLDPFQQ 120
```

Qy 121 IYGKRMGGLLFIPALMGEMFWAAAFSALGATISVIIDVDMHISVIISALIATLYTLVGG 180  
 Db 121 IYGKRMGGLLFIPALMGEMFWAAAFSALGATISVIIDVDVNISVIVSALIATLYTLVGG 180

Qy 181 LYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSSEVYSW 240  
 Db 181 LYSVAYTDVVQLFCIFIGLWISVPFALSHPAVTDIGFTAVHAKYQSPWLGTIESVEVYTW 240

Qy 241 LDSFLLMLLGGIPWQAYFQRVLSSSSATYAQVLSFLAAGCLVMAIPAILIGAIGASTDW 300  
 Db 241 LDNFLMLLGGIPWQAYFQRVLSSSSATYAQVLSFLAAGCLVMALPAICIGAIGASTDW 300

Qy 301 NQTAYGLPDPKTTEEADMILPIVLQYLCVPYISFFGLGAVSAAMSSADSSILSASSMFA 360  
 Db 301 NQTAYGFDPDKTKEEADMILPIVLQYLCVPYISFFGLGAVSAAMSSADSSILSASSMFA 360

Qy 361 RNIYQLSFRQNASDKEIVVMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQ 420  
 Db 361 RNIYQLSFRQNASDKEIVVMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIIIFPQ 420

Qy 421 LLCVLFVKGTNTYGAVAGYVSGFLRITGGEPYLYLQPLIFYPGYPPDDNGIYNQKFPFK 480  
 Db 421 LLCVLFVKGTNTYGAVAGYIFGLFLRITGGEPYLYLQPLIFYPGYPPDKNGIYNQKFPFK 480

Qy 481 TLMVTSFLTNICISYLAKYLFESGTLPPKLDVFDVAVRHSEENMDKTLVKNENIKLD 540  
 Db 481 TLMVTSFFTNICVSYLAKYLFESGTLPPKLDIFDAVVRHSEENMDKTLVRNENIKLN 540

Qy 541 ELALVKPRQSMTLSTFTNKEAFLDVDSSPEGSGTEDNLQ 580  
 Db 541 ELAPVKPRQSLTSLSTFTNKEALLDVDSSPEGSGTEDNLQ 580

RESULT 6

US-09-911-077A-4

; Sequence 4, Application US/09911077A

; Publication No. US20030114399A1

; GENERAL INFORMATION:

; APPLICANT: BLAKELY, RANDY D.

; APPLICANT: APPARSUNDARAM, SUBRAMANIAM

; APPLICANT: FERGUSON, SHAWN

; TITLE OF INVENTION: HUMAN AND MOUSE CHOLINE TRANSPORTER cDNA

; FILE REFERENCE: VBLT:008US

; CURRENT APPLICATION NUMBER: US/09/911,077A

; CURRENT FILING DATE: 2001-07-23

; NUMBER OF SEQ ID NOS: 27

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 4

; LENGTH: 580

; TYPE: PRT

; ORGANISM: Mus musculus

US-09-911-077A-4

Query Match 94.0%; Score 2795; DB 10; Length 580;

Best Local Similarity 92.6%; Pred. No. 9.8e-253;

Matches 537; Conservative 23; Mismatches 20; Indels 0; Gaps 0;



; LENGTH: 580  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-09-911-077A-24

Query Match 94.0%; Score 2795; DB 10; Length 580;  
Best Local Similarity 92.6%; Pred. No. 9.8e-253;  
Matches 537; Conservative 23; Mismatches 20; Indels 0; Gaps 0;

```
Qy      1 MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTA 60
      | | | | | : | | : | | | | | | | | | | | : | | | | : | | | | | | | | | | | |
Db      1 MPFHVEGLVAIILFYLLIFLVGIWAAWKTKNNGNPEERSEAIIVGGRDIGLLVGGFTMTA 60

Qy     61 TWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ 120
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | : |
Db     61 TWVGGGYINGTAEAVYVPGYGLAWAHAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFKQ 120

Qy    121 IYGKRMGGLLFIPALMGEMFWAAAI FSALGATISVIIDVDMHISVIISALIATLYTLVGG 180
      | | | | | | | | | | | | | | | | | | | | | | | | | | | : | | | | | | | |
Db    121 IYGKRMGGLLFIPALMGEMFWAAAI FSALGATISVIIDVDMNISVIVSALIAILYTLVGG 180

Qy    181 LYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSSEVYSW 240
      | | | | | | | | | | : | | | | | | | | | | | | | | | | | | | | | : | | | : |
Db    181 LYSVAYTDVVQLFCIFIGLWISVPFALSHPAVTDIGFTAVHAKYQSPWLGTIESVEVYTW 240

Qy    241 LDSFLLMLGGIPWQAYFQRLSSSSATYAQVLSFLAAFGLVMAIPAILIGAIGASTDW 300
      | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    241 LDNFLLMLGGIPWQAYFQRLSSSSATYAQVLSFLAAFGLVMALPAICIGAIGASTDW 300

Qy    301 NQTAYGLPDPKTTTEADMILPIVLQYLCVPYISFFGLGAVSAAVMSSADSSILSASSMFA 360
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    301 NQTAYGYDPDKTKEEADMILPIVLQYLCVPYISFFGLGAVSAAVMSSADSSILSASSMFA 360

Qy    361 RNIYQLSFRQNASDKEIVWVMRITVVFVGASATAMALLTKTVYGLWYLSSDLVYIVIFPQ 420
      | | | | | | | | | | | | | | | | | | | | | | | | | | | : | | | |
Db    361 RNIYQLSFRQNASDKEIVWVMRITVLVFGASATAMALLTKTVYGLWYLSSDLVYIIIFPQ 420

Qy    421 LLCVLFVKGTNTYGAVAGYVSGFLRITGGEPYLYLQPLIFYPGYPPDDNGIYNQKFPFK 480
      | | | | | : | | | | | | | | | : | | | | | | | | | | | | | | | | | | | |
Db    421 LLCVLFIKGTNTYGAVAGYIFGLFLRITGGEPYLYLQPLIFYPGYSDKNGIYNQRFPPK 480

Qy    481 TLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDVAVVARHSEENMDKTILVKNENIKLD 540
      | | : | | | | | | | | | : | | | | | | | | | | | | | | | | | | | : | | | |
Db    481 TLSMVTSTFTNICVSYLAKYLFESGTLPPKLDVFDVAVVARHSEENMDKTILVRNENIKLN 540

Qy    541 ELALVKPRQSMSTLSSTFTNKEAFLDVDSSPEGSGTEDNLQ 580
      | | | | | | | | : | | | | | | | | | | | | | | | | | | | | | |
Db    541 ELAPVKPRQSLTSLSTFTNKEALLDVDSSPEGSGTEDNLQ 580
```

RESULT 8  
US-10-241-784-2  
; Sequence 2, Application US/10241784  
; Publication No. US20040048261A1  
; GENERAL INFORMATION:  
; APPLICANT: Bayer Corporation





```
; Sequence 8, Application US/09911077A
; Publication No. US20030114399A1
; GENERAL INFORMATION:
; APPLICANT: BLAKELY, RANDY D.
; APPLICANT: APPARSUNDARAM, SUBRAMANIAM
; APPLICANT: FERGUSON, SHAWN
; TITLE OF INVENTION: HUMAN AND MOUSE CHOLINE TRANSPORTER cDNA
; FILE REFERENCE: VBLT:008US
; CURRENT APPLICATION NUMBER: US/09/911,077A
; CURRENT FILING DATE: 2001-07-23
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 576
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-911-077A-8
```

```
Query Match          48.9%; Score 1453; DB 10; Length 576;
Best Local Similarity 50.5%; Pred. No. 5e-127;
Matches 295; Conservative 95; Mismatches 150; Indels 44; Gaps 9;
```

```
Qy      7 GLIAIIVFYLLILLVGIWAAWRTKNSGSAEER---SEAIIVGGRDIGLLVGGFTMTATW 62
|::||: ||:||||:|||| |::| | |:: ||:| || ||||||
Db      6 GIVAIVFFYVLILVVGWIWAGRKSKESEAGAATEEVMLAGRNI GTLVGIFTMTATW 65

Qy     63 VGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQIY 122
||| |||||::| | |||:|::|::| || | |::| ||||| |
Db     66 VGGAYINGTAEALY--NGGLLGCQAPVGYAISLVMGGLLFAKKMREEGYITMLDPFQHKY 123

Qy    123 GKRMGGLLFIPALMGEMFWAAAFSALGATISVIIDVDMHISVIISALIATLYTLVGGLY 182
|::|::|::|::|::| || ||| |||||:|::|::|::|::| || | |::| || |
Db    124 GQRIGGLMYVPALLGETFWTAAILSALGATLSVILGIDMNASVTL SACIAVFYTF TGGYY 183

Qy    183 SVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVD S-SEVYSWL 241
:|||||:|||||: || |::| | | | |::|::| | |::|
Db    184 AVAYTDVVQLFCIFVGLWVCVPAAMVHDGAKDISRNA-----GDWIGEIGGFKETSLWI 237

Qy    242 DSFLLMLLGGIPWQAYFQRVLSSSSATYAQVLSFLAAGCLVMAIPAILIGAIGASTDWN 301
| |||: ||||| |||||::| || |||:| ||:| || | ||| :||
Db    238 DCMLLLVFGGIPWQVYFQRVLSSKTAHGAQTLSEFVAGVGCILMAIPPALIGAIARNTDWR 297

Qy    302 QTAYGLPDPKTTEEA-----DMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSA 355
| | : | | : |::|:| || |::| ||||| |||||:| ||
Db    298 MTDYSPWNNGTKVESIPDKRNMVPLVLFQYLTPRWVAFI GLGAVSAAVMSSADSSVL SA 357

Qy    356 SSMFARNIYQLSFRQNASDKEIVVMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYI 415
:|||| ||::|: |::|::| ||| : | || | |::| |||| :||||
Db    358 ASMFAHNIWKL TIRPHASEKEVIIVMRIAII CVGIMATIMALT IQSIYGLWYLCADLVYV 417

Qy    416 VIFPQLLCVLFVKGTNTYGA VAGYVSGFLRITGGEPYLYLQPLIFYPGYYPDDNGIYNQ 475
::| |||||::: :|||::| || |||: ||| : | | | :|: |
Db    418 ILFPQLLCVVYMPRSNTYGS LAGYAVGLVLR LIGGEPLVSLPAFFHYPMY---TDGV--Q 472

Qy    476 KFPFKTLAMVTSFLT NICISYLAKYLFESGTLPPKLDVFDVAV---ARHSEENMDKTILV 532
|||:| ||::| | :| :: ||:| | |::| | | | | : |
Db    473 YFPFRTTAMLSSMATIYIVSIQSEKLFKSGRLSPEWDMGCVVNIPI DHVPLPSDVSFAV 532
```

Qy 533 KNENIKL-----DELALVKPRQSM TLSSTFTN 559  
 :| : : || | : | : || :|  
 Db 533 SSETLNMKAPNGTPAPVHPNQPSDENTLLHPYSYDQSYSTNSN 576

RESULT 10

US-09-733-630-2

; Sequence 2, Application US/09733630  
 ; Patent No. US20020034799A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Donoho, Gregory  
 ; APPLICANT: Scoville, John  
 ; APPLICANT: Turner, C. Alexander Jr.  
 ; APPLICANT: Freidrich, Glenn  
 ; APPLICANT: Zambrowicz, Brian  
 ; APPLICANT: Abuin, Alejandro  
 ; APPLICANT: Sands, Arthur T.  
 ; TITLE OF INVENTION: No. US20020034799A1el Human Transporter Protein and  
 ; TITLE OF INVENTION: Polynucleotides Encoding the Same  
 ; FILE REFERENCE: LEX-0106-USA  
 ; CURRENT APPLICATION NUMBER: US/09/733,630  
 ; CURRENT FILING DATE: 2000-12-08  
 ; PRIOR APPLICATION NUMBER: US 60/170,137  
 ; PRIOR FILING DATE: 1999-12-10  
 ; NUMBER OF SEQ ID NOS: 3  
 ; SOFTWARE: FastSEQ for Windows Version 4.0  
 ; SEQ ID NO 2  
 ; LENGTH: 675  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-09-733-630-2

Query Match 10.5%; Score 311.5; DB 9; Length 675;  
 Best Local Similarity 23.0%; Pred. No. 5.3e-20;  
 Matches 152; Conservative 112; Mismatches 235; Indels 161; Gaps 28;

Qy 2 AFHVEGL----IAIIVFY-LLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGF 56  
 || :|| ||:| | | :| ||:| :|| : : : | | : |  
 Db 18 AFPQKGLEPGDIAVLVLYFLFVLAVGLWSTVKTK-----RDTVKG YFLAGGDMVWWPVGA 72  
 Qy 57 TMTATWVG GGYINGTAEAVYVPGYGLAWAQAPIGYSLSLI-LGGLF-----FAKPMRS 108  
 :: | : || | : ||| : | | | : | || | :  
 Db 73 SLFASNVGSGHF-----IGLAGSSAATGISVSAYELNGLFSVIMLAWIFLPIYI 121  
 Qy 109 KGYVTMLDPPFQQIYGKRMGGLLFIPALMGEMFWAAAI FSAL-----GAT-ISVIIDVDM 161  
 | || : : : || || : || : : || : : || | : :|  
 Db 122 AGQVTTMPEYLR--KRFGGIR-IPILAVLYLFYIYFTKISVDMYAGAIFIQQSLHLDL 177  
 Qy 162 HISVIISALIATLYTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVH 221  
 : : : : | : || : || : | || : | : : | : : | || | :  
 Db 178 YLAIVGLLAI TAVYT VAGGLAAVIYTDALQTLIMLIGALTLMGY--SFAAVG--GMEGLK 233  
 Qy 222 AKY-----QKPWLGT VDSSEVYS-WLDSFLLML 249  
 || || | : : | |  
 Db 234 EKYFLALASNRSENSSCGLPREDAFHIFRDPLTSDLPWPGVLF GMSIPSLWY----- 285

QY 250 GGIPW---QAYFQRVLSSSSATYAQVLSFLAAGCLVMAIPAILIGAIGASTDWNQTAYG 306  
 | | | | : : : : : : : : : : : : : : : : : :  
 Db 286 ----WCTDQVIVQRTLAAKNLSHAKGGALMAAYLKVLPLFIMVFPGMVSRILFPDQVA-- 339  
 QY 307 LPDPKTTEE-----ADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSM 358  
 ||: : : : | : : | : : : : ||: || | ||: :  
 Db 340 CADPEICQKICSNPSGCSDIAYPKLVLELLPTGLRGLMMAVMVAALMSSLTSIFNSASTI 399  
 QY 359 FARNIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKTVYGLW----- 406  
 | : : | | : | : | : | : | : | : | : | : | :  
 Db 400 FTMDLWN-HLRPRASEKELMIVGRVFFV-----LLVLVLSILWIPVVQASQGGQL 447  
 QY 407 --YLSSDLVYI-----VIFPQLLCVLFVKGTNTYGAVAGYVSGFLRITG-GEPLYLYLQP 458  
 | : | : | : | : | : | : | : | : | : | : | :  
 Db 448 FIYIQSISSYLQPPVAVVF---IMGCFWKRTNEKGAFWGLISGLLLGLVRLVLDIFIYVQP 504  
 QY 459 LIFYPGYYPDDNGIYNQKFFPKTLAMVTSFLTNICISYLAKYLFESGTLPPKLDV----- 513  
 ||:  
 Db 505 RC----DQPDERPVLVKSIHYLYFSMILSTVTLITVSTVSWF-----TEPPSKEMVSHLT 555  
 QY 514 ----FDAVVARHSEENMDKTILVKNNENIKLD-----ELALVKPRQSMTLSSTFTNKEA 562  
 | | : | : : : : : : : : : : | | | | : :  
 Db 556 WFTRHDPVVQKEQAPPAAPLSLTLSQNGMPEASSSSSVQFEMVQENTSKTHSCDMTPKQS 615

RESULT 11  
 US-10-156-761-12818  
 ; Sequence 12818, Application US/10156761  
 ; Publication No. US20030119018A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: OMURA, SATOSHI  
 ; APPLICANT: IKEDA, HARUO  
 ; APPLICANT: ISHIKAWA, JUN  
 ; APPLICANT: HORIKAWA, HIROSHI  
 ; APPLICANT: SHIBA, TADAYOSHI  
 ; APPLICANT: SAKAKI, YOSHIYUKI  
 ; APPLICANT: HATTORI, MASAHIRA  
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
 ; FILE REFERENCE: 249-262  
 ; CURRENT APPLICATION NUMBER: US/10/156,761  
 ; CURRENT FILING DATE: 2002-05-29  
 ; PRIOR APPLICATION NUMBER: JP 2001-204089  
 ; PRIOR FILING DATE: 2001-05-30  
 ; PRIOR APPLICATION NUMBER: JP 2001-272697  
 ; PRIOR FILING DATE: 2001-08-02  
 ; NUMBER OF SEQ ID NOS: 15109  
 ; SEQ ID NO 12818  
 ; LENGTH: 486  
 ; TYPE: PRT  
 ; ORGANISM: Streptomyces avermitilis  
 US-10-156-761-12818

Query Match 10.3%; Score 306; DB 14; Length 486;  
 Best Local Similarity 27.7%; Pred. No. 1.1e-19;  
 Matches 130; Conservative 84; Mismatches 194; Indels 62; Gaps 20;

QY 11 IIVFYLL-ILLVGIWAAWRTKNSGSAERSEAIIVGGRDIGLLVGGFTMTATWVGGGYIN 69

```

      :|| || :| :| | | | :      :| || :| : || | :||
Db      7 VIVVYLAGMLAMGWWGMRRRAKSKSD-----FLVAGRRLGPMYSGTMAAIVLGGASTI 59

Qy      70 GTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQIYGKRMGGL 129
      | | | | | | | | | | | | : | | | | : : | | | |
Db      60 GGVGLGYKYGLSGAWMVFAIGLGL-LALSVFFSARIARLKVY-TVSEMLDLRYGGRAG-- 115

Qy      130 LFIPALMGEMFWAAAFSALGATIS-----VIIDVDMHISVIISALIATLYTLVGGLYS 183
      : | : || : | : || : | : || : | : || : | : || : |
Db      116 ----VISGVVMWAYTLMIAVTSTIAYATIFDVLFDNMRTLAILGGSIVVAYSTLGGMWS 171

Qy      184 VAYTDVVQLFCIFVG-LWISVPFALSHPAVADIGFTAVHAK----YQKPWLGTVDSSEVY 238
      : ||: || : | | : :| | : | ||: || | | | | : :
Db      172 ITLTDVMVQFVVKTIQVLLLLLPIAI----VKAGGFSAMKAKLPTEYFDP-LG-IGGETIF 225

Qy      239 SWLDSFLLMLGGIPWQAYFQVRLSSSSATYAQVLSFLAAFGCLVMAIPAILIGAIGAST 298
      :: : | : | : | : || : : | | : : | || | : : ||
Db      226 TYV---LIYTFGMLIGQDIWQVFTARSDDTAKWGGTVAGTYCLVYALAGAVIG----- 276

Qy      299 DWNQTAYGLPDKTTEEADMILPIVLQYLCVPYISFFGLGAVSAAVMSSADSSILSASSM 358
      || : | | || : : || : | | |||| : : : : : :
Db      277 ----TAAKVLYP-TLPSADSAFATIVKDELPGVVRGLVLAALAAVMSTSSGALIACATV 331

Qy      359 FARNIYQ----LSFRQNASDKEIVWVMRITVFVFGASATAMAL-LTKTVYGLWYLSSDLV 413
      :| : :| | : | | : : | | : | : | | | : ||
Db      332 ANNDIWSRLRGVSSRK-GDDHDEVGRNRLFILVMGVAVICTAIALNDVVEALTVAYNLLV 390

Qy      414 YIVIFPQLLCVLFVKGNTNTYGAVAGYVSG----LFLRITGG----EPYLY 455
      : : | | :| : || : ||: | : | : | | | | || |
Db      391 GGLLVPILGGLLWKRGT-VHGALASVIVGGLAVIGLMATFGILANEPVYY 439

```

# RESULT 12

US-10-119-988-12

; Sequence 12, Application US/10119988

; Publication No. US20030054453A1

## ; GENERAL INFORMATION:

; APPLICANT: Curtis, Rory A.J.

; APPLICANT: Chen, Hong

; APPLICANT: Millennium Pharmaceuticals Inc.

; TITLE OF INVENTION: 68723, Sodium/Glucose Cotransporter

; TITLE OF INVENTION: Family Members and Uses Therefor

; FILE REFERENCE: MPI01-103P1RNM

; CURRENT APPLICATION NUMBER: US/10/119,988

; CURRENT FILING DATE: 2002-04-10

; PRIOR APPLICATION NUMBER: 60/282,764

; PRIOR FILING DATE: 2001-04-10

; NUMBER OF SEQ ID NOS: 18

; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 12

; LENGTH: 664

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-119-988-12

Query Match 10.3%; Score 306; DB 14; Length 664;

Best Local Similarity 22.8%; Pred. No. 1.7e-19;

Matches	148;	Conservative	104;	Mismatches	218;	Indels	178;	Gaps	30;
Qy	11	IIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGMTMTATWVGGGYING	70						
Db	32	IVIYFVVVMAVGLWAMFST-NRGTV----GGFFLAGRSMVWVPIGASLFASNIGSGHFVG	86						
Qy	71	TAEAVYVPGYGLAWAQAPIGYS-----LSLILGGLFFAKPMRSK-GYVTMLDPFQQIYGK	124						
Db	87	LA-----GTGAASGIAIGGFENALVLVVVLGWLFV--PIYIKAGVVTM----PEYLRK	134						
Qy	125	RMGG-----LLFIPALMGEMFWAAAFSALGATISVIIDVDMHISVIISALIA	172						
Db	135	RFGGQRIQVYLSLLSLLLYIFTKISADIFSGAIF-----INLALGLNLYLAIFLLLAIT	188						
Qy	173	TLYTLVGGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQK--PWL-	229						
Db	189	ALYTITGGLAAVIYTDTLQTVIMLVGSLILTGFAPHEVG---GYDAFMEKYMKAIPTIV	244						
Qy	230	---GTVDSSEVYS-WLDSFLLL--MLGGIPW-----QAYFQRVLSS	264						
Db	245	SDGNTTFQEKCYTPRADSFHIFRDPLTGDLPWPGFIFGMSILTLWYWCTDQVIVQRCLSA	304						
Qy	265	SSATYAQ----VLSFLAAFGCLVMAIPAIL-----IGAI-----GASTDWNQT	303						
Db	305	KNMSHVKGGCILCGYLKMPMFIMVMPGMISRILYTEKIACVVPSECEKYCGTKVGCTNI	364						
Qy	304	AYGLPDPKTTTEADMILPIVLQYLCVPYISFFGLGAVSAAVMSSADSSILSASSMFARNI	363						
Db	365	AY-----PTLVVELMPNGLRGLMLSVMLASIMSSLTSIFNSASTLFTMDI	409						
Qy	364	YQLSFRQNASDKEIVVMRITVFV-FGASATAMALLTKTVYG--LWYLSSDLVYI--VIF	418						
Db	410	Y-AKVRKRASEKELMIAGRLFILVLIGISIAWVPVIVQSAQSGQLFDYIQSITSYLGPPIA	468						
Qy	419	PQLLCVLFVKGTNTYGAVAGYVSGFLRLI-----TG-----GEPYLY	455						
Db	469	AVFLLAIFWKRNVNEPGAFWGLILGLLIGISRMITEFAYGTGSCMEPSNCPTIICGVHYLY	528						
Qy	456	LQPLIFYPGYYPDDNGIYNQKFPFKTLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFD	515						
Db	529	FAILLF-----AISFITIVVISLLTKPI-----PDVHLYR	558						
Qy	516	AV--VARHSEENMDKTIIVKNENIKLDELALVKPRQSMTLSSTFTNKE	561						
Db	559	LCWSLRNSKEERID--LDAEEENIQ-----EGPKETIEIETQVPEKK	598						

```
; CURRENT APPLICATION NUMBER: US/09/928,530
; CURRENT FILING DATE: 2001-08-13
; PRIOR APPLICATION NUMBER: 60/227,068
; PRIOR FILING DATE: 2000-08-22
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 675
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-928-530-2
```

```
Query Match          10.0%; Score 298.5; DB 9; Length 675;
Best Local Similarity 22.7%; Pred. No. 8.7e-19;
Matches 149; Conservative 115; Mismatches 238; Indels 155; Gaps 29;
```

```
Qy      2 AFHVEGL----IAIIVFY-LLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGF 56
      || :||  ||::| | | :| ||:: :|| : : | : ||
Db     18 AFPQKGLEPGDIAVLVLYFLFVLAVGLWSTVTKR----DTVKGYFLAEGNMVWWPVGA- 72

Qy     57 TMTATWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLS-----LILGGLFFAKPMRSKGY 111
      :: |: || |: | | | | : | |: | |: | |: |
Db     73 SLFASNVGSGHFIGLA-----GSGAATGISVSAYELNGLFSVLMLAWIFL--PIYIAGQ 124

Qy    112 VTMLDPFQQIYGKRMGGLLFIPALMGEMFWAAAFSALGATI---SVIDID---VDMHIS 164
      || : : : || ||: || : : ||: : : : | : : :
Db    125 VTTMPEYLR---KRFGGIR-IPITLAVLYLFYIFTKISVDMYAGAFIQQSSHLDLILA 180

Qy    165 VIISALIATLYTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKY 224
      :: | : ||: || | || | : | : | : | : | : |
Db    181 IVGLLAITAVYTVAGGLAAVIYTDALQTLIMLIGALTLMGY--SFAAVG--GMEGLKEKY 236

Qy    225 -----QKPWLGTVDSSEVYS-WLDSFLLMLGGI 252
      || | : : | |
Db    237 FLALASNRSENSSCGLPREDAFHIFRDPLTSDLPWPGVLFGMSIPSLWY----- 285

Qy    253 PW---QAYFQRVLSSSSATYAQVLSFLAAFGLVMAIPAILIGAIGASTDWNQTAYGLPD 309
      | | | || : : : : : : : : : : | : : : | |
Db    286 -WCTDQVIVQRTLAAKNLSHAKGGALMAAYLKVLPLFIMVFPGMVSRILFPDQVA--CAD 342

Qy    310 PKTTEE-----ADMILPIVLQYLCPVYISFFGLGAVSAAMSSADSSILSASSMFAR 361
      |: : : : |: | : : | | : : : ||: || | ||: :|
Db    343 PEICQKICSNPSGCSDIAYPKLVLELLPTGLRGLMMAVMVAALMSSLTSIFNSASTIFTM 402

Qy    362 NIYQLSFRQNASDKEIVWVMRITVVFVGASATAMALLTKTVYGLW-----Y 407
      :: : | ||: ||: | | : | || | || |
Db    403 DLWN-HLRPRASEKELMIVGRVFN-----LLLVLVSILWIPVVQASQGGQLFIY 450

Qy    408 LSSDLVYI-----VIFPQLLCVLFVKGTNTYGAVAGYVSGFLRITG-GEPLYLYQLPLIF 461
      : | |: | : | : | | || || | : || | : : :|
Db    451 IQSISSYLQPPVAVVF---IMGCFWKRTNEKGAFWGLISGLLLGLVRLVLDIFIYVQPRC- 506

Qy    462 YPGYYPDDNGIYNQKFPFKTLAMVTSFLTNICISYLAKYLFESGTLPPKLDV----- 513
      ||: : : : : |: | : | :| : : : | || :
Db    507 ---DQPDERPVLVKSIHYLYFSMILSTVTLITVSTVSWF-----TEPPSKEMVSHLTWFT 558

Qy    514 -FDAVVARHSEENMDKTILVKNENIKLD-----ELALVKPRQSMTLSSSTFTNKEA 562
```

Db                    | || :                    | :| :                    : :| :                    | | |                    | | : :  
559 RHD PVVQKEQAPPAAPLSLTLSQNGMPEASSSSSVQFEMVQENTSKTHSCDMTPKQS 615

RESULT 14

US-10-162-012-27

; Sequence 27, Application US/10162012  
; Publication No. US20030051660A1  
; GENERAL INFORMATION:  
; APPLICANT: Curtis, Rory A.J.  
; APPLICANT: Silos-Santiago, Inmaculada  
; APPLICANT: Gu, Wei  
; TITLE OF INVENTION: NOVEL HUMAN ION CHANNEL AND TRANSPORTER FAMILY MEMBERS  
; FILE REFERENCE: 10448-190001  
; CURRENT APPLICATION NUMBER: US/10/162,012  
; CURRENT FILING DATE: 2002-06-04  
; PRIOR APPLICATION NUMBER: US 60/209,845  
; PRIOR FILING DATE: 2000-06-06  
; PRIOR APPLICATION NUMBER: US 09/875,321  
; PRIOR FILING DATE: 2001-06-06  
; PRIOR APPLICATION NUMBER: PCT/US01/18340  
; PRIOR FILING DATE: 2001-06-06  
; PRIOR APPLICATION NUMBER: US 60/209,257  
; PRIOR FILING DATE: 2000-06-05  
; PRIOR APPLICATION NUMBER: US 09/875,423  
; PRIOR FILING DATE: 2001-06-05  
; PRIOR APPLICATION NUMBER: PCT/US01/18398  
; PRIOR FILING DATE: 2001-06-05  
; PRIOR APPLICATION NUMBER: US 60/209,238  
; PRIOR FILING DATE: 2000-06-05  
; PRIOR APPLICATION NUMBER: US 09/875,363  
; PRIOR FILING DATE: 2001-06-05  
; PRIOR APPLICATION NUMBER: PCT/US01/18247  
; PRIOR FILING DATE: 2001-06-05  
; PRIOR APPLICATION NUMBER: US 60/227,068  
; PRIOR FILING DATE: 2000-08-22  
; PRIOR APPLICATION NUMBER: US 09/928,530  
; PRIOR FILING DATE: 2001-08-13  
; PRIOR APPLICATION NUMBER: PCT/US01/25475  
; PRIOR FILING DATE: 2001-08-15  
; PRIOR APPLICATION NUMBER: US 60/226,770  
; PRIOR FILING DATE: 2000-08-21  
; PRIOR APPLICATION NUMBER: US 09/934,421  
; PRIOR FILING DATE: 2001-08-21  
; PRIOR APPLICATION NUMBER: PCT/US01/26096  
; PRIOR FILING DATE: 2001-08-21  
; PRIOR APPLICATION NUMBER: US 60/279,281  
; PRIOR FILING DATE: 2001-03-28  
; PRIOR APPLICATION NUMBER: US 10/109,029  
; PRIOR FILING DATE: 2002-03-28  
; PRIOR APPLICATION NUMBER: PCT/US02/09728  
; PRIOR FILING DATE: 2002-03-28  
; PRIOR APPLICATION NUMBER: US 60/290,288  
; PRIOR FILING DATE: 2001-05-11  
; PRIOR APPLICATION NUMBER: US (not assigned)  
; PRIOR FILING DATE: 2002-05-13  
; NUMBER OF SEQ ID NOS: 48

; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 27  
; LENGTH: 675  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-162-012-27

Query Match 10.0%; Score 298.5; DB 14; Length 675;  
Best Local Similarity 22.7%; Pred. No. 8.7e-19;  
Matches 149; Conservative 115; Mismatches 238; Indels 155; Gaps 29;

```
Qy      2 AFHVEGL----IAIIVFY-LLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGF 56
      || :||  ||::| | | :| ||::| :|| : : | : ||
Db      18 AFPQKGLEPGDIAVLVLYFLFVLAVGLWSTVKTKR----DTVKGYFLAEGNMVWWPVGA- 72

Qy      57 TMTATWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLS-----LILGGLFFAKPMRSKY 111
      :: | : || | : | | | : | : | : | : | : | : |
Db      73 SLFASNVGSGHFIGLA-----GSGAATGISVSAYELNGLFSVLMLAWIFL--PIYIAGQ 124

Qy     112 VTMLDPFQQIYGKRMGGLLFIPALMGEMFWAAAIFSAIGATI---SVID--VDMHIS 164
      || : : : || ||: || :: : || : : : : | : : :
Db     125 VTTMPEYLR---KREGGIR-IPILAVLYLFYIFTKISVDMYAGAFIQSSHLDLILA 180

Qy     165 VIISALIATLYTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKY 224
      :: | : ||: ||| :| ||| :| : :| : : | || | : ||
Db     181 IVGLLAITAVYTVAGGLAAVIYTDALQTLIMLIGALTIMGY--SFAAVG--GMEGLKEKY 236

Qy     225 -----QKPWLGTVDSSSEVYS-WLDSFLLLMLGGI 252
      || | : : ||
Db     237 FLALASNRSENSSCGLPREDAFHIRDPLTSDLPPWPGVLFGMSIPSLWY----- 285

Qy     253 PW---QAYFQRLSSSSATYAQVLSFLAAGCLVMAIPAILIGAIGASTDWNQTAYGLPD 309
      | | | || : : : || : : || : : : | : : : | | |
Db     286 -WCTDQVIVQRTLAAKNLSHAKGGALMAAYLKVLPLFIMVFPGMVSRILFPDQVA--CAD 342

Qy     310 PKTTEE-----ADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFAR 361
      | : : : : || : | : | : : : || : || | | || : |
Db     343 PEICQKICSNPSCSDIAYPKLVLELLPTGLRGLMMAMVMAALMSSSLTSIFNSASTIFTM 402

Qy     362 NIYQLSFRQNASDKEIVWVMRITVVFVGASATAMALLTKTVYGLW-----Y 407
      :: : | ||: || : | : | || | || |
Db     403 DLWN-HLRPRASEKELMIVGRVFV-----LLLVLVSILWIPVVQASQGGQLFIY 450

Qy     408 LSSDLVYI-----VIFPQLLCVLFVKGTNTYGAVAGYVSGFLRITG-GEPLYLQPLIF 461
      : | | : : | : | | || || | : || | : : || : ||
Db     451 IQSISSYLQPPVAVVF---IMGCFWKRTNEKGAFWGLISGLLLGLVRLVLDIFIYVQPRC- 506

Qy     462 YPGYYPDDNGIYNQKFPFKTLAMVTSFLTNICISYLAKYLFESGTLPPKLDV----- 513
      || : : : : : || : | : | : | : : | || : :
Db     507 ---DQDERPVLVKSIHYLYFSMILSTVTLITVSTVSWF-----TEPPSKEMVSHLTWFT 558

Qy     514 -FDAVVARHSEENMDKTILVKNENIKLD-----ELALVKPRQSMTLSSTFTNKEA 562
      | || : | : || : : : : | : | | | | : :
Db     559 RHDPVVQKEQAPPAAPLSLTLSQNGMPEASSSSSVQFEMVQENTSKTHSCDMPKQS 615
```

RESULT 15



US-10-162-102-27  
; Sequence 27, Application US/10162102  
; Publication No. US20030232336A1  
; GENERAL INFORMATION:  
; APPLICANT: Curtis, Rory A.J.  
; APPLICANT: Silos-Santiago, Inmaculada  
; APPLICANT: Gu, Wei  
; TITLE OF INVENTION: NOVEL HUMAN ION CHANNEL AND TRANSPORTER FAMILY MEMBERS  
; FILE REFERENCE: 10448-190001  
; CURRENT APPLICATION NUMBER: US/10/162,102  
; CURRENT FILING DATE: 2003-04-04  
; PRIOR APPLICATION NUMBER: US 60/209,845  
; PRIOR FILING DATE: 2000-06-06  
; PRIOR APPLICATION NUMBER: US 09/875,321  
; PRIOR FILING DATE: 2001-06-06  
; PRIOR APPLICATION NUMBER: PCT/US01/18340  
; PRIOR FILING DATE: 2001-06-06  
; PRIOR APPLICATION NUMBER: US 60/209,257  
; PRIOR FILING DATE: 2000-06-05  
; PRIOR APPLICATION NUMBER: US 09/875,423  
; PRIOR FILING DATE: 2001-06-05  
; PRIOR APPLICATION NUMBER: PCT/US01/18398  
; PRIOR FILING DATE: 2001-06-05  
; PRIOR APPLICATION NUMBER: US 60/209,238  
; PRIOR FILING DATE: 2000-06-05  
; PRIOR APPLICATION NUMBER: US 09/875,363  
; PRIOR FILING DATE: 2001-06-05  
; PRIOR APPLICATION NUMBER: PCT/US01/18247  
; PRIOR FILING DATE: 2001-06-05  
; PRIOR APPLICATION NUMBER: US 60/227,068  
; PRIOR FILING DATE: 2000-08-22  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 48  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 27  
; LENGTH: 675  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-162-102-27

Query Match 10.0%; Score 298.5; DB 15; Length 675;  
Best Local Similarity 22.7%; Pred. No. 8.7e-19;  
Matches 149; Conservative 115; Mismatches 238; Indels 155; Gaps 29;

Qy 2 AFHVEGL----IAIIVFY-LLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGF 56  
|| :|| ||:| | | :| ||:| :|| : : | : ||  
Db 18 AFPQKGLEPGDIAVLVLYFLFVLAVGLWSTVKTKR----DTVKGYFLAEGNMVWWPVGA- 72  
Qy 57 TMTATWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLS-----LILGGLFFAKPMRSKGY 111  
:: | : || | : | | | : | | : | : | : |  
Db 73 SLFASNVGSGHFIGLA-----GSGAATGISVSAYELNGLFSVLMLAWIFL--PIYIAGQ 124  
Qy 112 VTMLDPFQQIYGKRMGGLLFIPALMGEMFWAAAI FSALGATI---SVIID----VDMHIS 164  
|| : : : || ||: || : : : || : : : : | : : : :  
Db 125 VTTMPEYLR---KRFGGIR-IPILAVLYLFIYIFTKISVDMYAGAIFIQQSSHLDLYLA 180  
Qy 165 VIISALIATLYTLVGGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKY 224

```

      :: | :||: ||| :| ||| :| : : | | | : ||
Db    181 IVGLLAITAVYTVAGGLAAVIYTDALQTLIMLIGALTLMGY--SFAAVG--GMEGLKEY 236
Qy    225 -----QKPWLGTVDSSSEVYS-WLDSFLLMLGGI 252
      || | : : | |
Db    237 FLALASNRSENSSCGLPREDAFHIFRDPLTSDLPWPGVLFMGMSIPSLWY----- 285
Qy    253 PW---QAYFQRVLSSSSATYAQVLSFLAAFGLVMAIPAILIGAIGASTDWNQTAYGLPD 309,
      | | || |:: : ::| : :||: :: : | : :| | |
Db    286 -WCTDQVIVQRTLAAKNLSHAKGGALMAAYLKVLPLFIMVFGMVSRILFPDQVA--CAD 342
Qy    310 PKTTEE-----ADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFAR 361
      |: :: :||: | :: | | : : : ||: || | ||::|
Db    343 PEICQKICSNPSGCSDIAYPKLVLELLPTGLRGLMMAMVMAALMSSLTSIFNSASTIFTM 402
Qy    362 NIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKTVYGLW-----Y 407
      :: | ||: ||: | | | || | || |
Db    403 DLWN-HLRPRASEKELMIVGRVFV-----LLLVLVSILWIPVVQASQGGQLFIY 450
Qy    408 LSSDLVYI-----VIFPQLLCVLFVKGTNTYGAVAGYVSGFLRITG-GEPLYLQPLIF 461
      : | | : | : | | || || | : || | : :||: ||
Db    451 IQSISSYLQPPVAVVF---IMGCFWKRTNEKGAFWGLISGLLLGLVRLVLDIFIYVQPRC- 506
Qy    462 YPGYYPDDNGIYNQKFPFKTLAMVTSFLTNICISYLAKYLFESGTLPPKLDV----- 513
      ||: : : : :| : | : | : : | || ::
Db    507 ---DQDERPVLVKSIIHYLYFSMILSTVTLITVSTVSWF-----TEPPSKEMVSHLTWFT 558
Qy    514 -FDAVVARHSEENMDKTILVKNENIKLD-----ELALVKPRQSMTLSSTFTNKEA 562
      | || : | :| : : : :| : | | | | : :
Db    559 RHDPVVQKEQAPPAAPLSLTLSQNGMPEASSSSSVQFEMVQENTSKTHSCDMTPKQS 615

```

Search completed: March 22, 2004, 15:45:44  
Job time : 536 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 22, 2004, 15:17:05 ; Search time 105 Seconds  
(without alignments)  
1742.862 Million cell updates/sec

Title: US-10-069-541-6  
Perfect score: 2972  
Sequence: 1 MAFHVEGLIAIIVFYLLILL.....EAFLDVDSSPEGSGTEDNLQ 580

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_25:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archeap:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Score	Query	Match	Length	DB	ID	Description
No.							

---

1	2972	100.0	580	4	Q9GZV3	Q9gzv3 homo sapien
2	2820	94.9	580	11	Q9JMD7	Q9jmd7 rattus norv
3	2806	94.4	580	11	Q8BGY9	Q8bgy9 mus musculu
4	2795	94.0	580	11	Q99PK3	Q99pk3 mus musculu
5	2791	93.9	580	11	Q9ESW5	Q9esw5 mus musculu
6	2253	75.8	584	13	Q8UWF0	Q8uwf0 torpedo mar
7	1679	56.5	377	13	Q8AV27	Q8av27 gallus gall
8	1557.5	52.4	614	5	Q9VE46	Q9ve46 drosophila
9	1530.5	51.5	579	5	Q9GPB1	Q9gpb1 limulus pol
10	1453	48.9	576	5	O02228	O02228 caenorhabdi
11	422.5	14.2	484	16	Q7UFM6	Q7ufm6 rhodopirell
12	405.5	13.6	462	16	Q8EXG7	Q8exg7 leptospira
13	381.5	12.8	479	16	Q8Y273	Q8y273 ralstonia s
14	344	11.6	492	17	Q9V2P3	Q9v2p3 pyrococcus
15	334	11.2	492	16	Q81Y52	Q81y52 bacillus an
16	333	11.2	493	16	Q81AD3	Q81ad3 bacillus ce
17	316	10.6	493	17	Q8U3M8	Q8u3m8 pyrococcus
18	314.5	10.6	665	11	Q9QXI6	Q9qxi6 mus musculu
19	314.5	10.6	665	11	Q8C3K6	Q8c3k6 mus musculu
20	312.5	10.5	665	11	Q9QXX5	Q9qxx5 mus musculu
21	311.5	10.5	480	16	Q8ERF0	Q8erf0 oceanobacil
22	311	10.5	665	11	Q8CCA7	Q8cca7 mus musculu
23	310.5	10.4	675	4	Q8WWX8	Q8wwx8 homo sapien
24	309	10.4	670	11	Q923I7	Q923i7 mus musculu
25	308.5	10.4	675	4	Q86Y55	Q86y55 homo sapien
26	307.5	10.3	675	4	Q96PP5	Q96pp5 homo sapien
27	306	10.3	486	16	Q82CQ9	Q82cq9 streptomyce
28	304	10.2	567	16	Q8EQQ5	Q8eqq5 oceanobacil
29	301	10.1	463	16	Q9I3S6	Q9i3s6 pseudomonas
30	300	10.1	500	16	Q9CN55	Q9cn55 pasteurilla
31	299.5	10.1	507	16	Q9K9E2	Q9k9e2 bacillus ha
32	299.5	10.1	546	16	Q8G6N6	Q8g6n6 bifidobacte
33	299	10.1	662	6	Q9BDF6	Q9bdf6 equus cabal
34	299	10.1	718	11	Q80WA5	Q80wa5 rattus norv
35	297.5	10.0	673	11	Q8K0E3	Q8k0e3 mus musculu
36	292	9.8	698	4	Q8WY15	Q8wyl5 homo sapien
37	291.5	9.8	535	16	Q8FQ71	Q8fq71 corynebacte
38	291	9.8	514	17	Q8TUR0	Q8tur0 methanosarc
39	290.5	9.8	674	6	Q863B5	Q863b5 oryctolagus
40	287	9.7	664	6	Q8MKB7	Q8mkb7 bos taurus
41	287	9.7	678	11	Q8VDT1	Q8vdt1 mus musculu
42	286.5	9.6	674	6	Q28728	Q28728 oryctolagus
43	286	9.6	491	17	O58753	O58753 pyrococcus
44	286	9.6	685	11	Q8BZW1	Q8bzw1 mus musculu
45	286	9.6	685	11	Q8BGU9	Q8bgu9 mus musculu

#### ALIGNMENTS

##### RESULT 1

Q9GZV3

ID Q9GZV3 PRELIMINARY; PRT; 580 AA.

AC Q9GZV3;

DT 01-MAR-2001 (TrEMBLrel. 16, Created)

DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)

DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

DE High affinity choline transporter (High-affinity choline transporter  
 DE CHT1).  
 GN CHT1 OR SLC5A7.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Hypothalamus;  
 RA Bruess M.;  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Hypothalamus;  
 RA Wieland A., Bonisch H., Bruss M.;  
 RT "Molecular cloning of the human and murine high affinity choline  
 RT transporters and characterization of the human gene-structure.";  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20483599; PubMed=11027560;  
 RA Apparsundaram S., Ferguson S.M., George A.L. Jr., Blakely R.D.;  
 RT "Molecular cloning of a human, hemicholinium-3-sensitive choline  
 RT transporter.";  
 RL Biochem. Biophys. Res. Commun. 276:862-867(2000).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Spinal cord;  
 RX PubMed=11068039;  
 RA Okuda T., Haga T.;  
 RT "Functional characterization of the human high-affinity choline  
 RT transporter.";  
 RL FEBS Lett. 484:92-97(2000).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA Bruess M.;  
 RL Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RA Wieland A., Bonisch H., Bruss M.;  
 RT "Molecular cloning of the human and murine high affinity choline  
 RT transporters and characterization of the human gene structure.";  
 RL Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AJ401466; CAC03717.1; -.  
 DR EMBL; AF276871; AAG25940.1; -.  
 DR EMBL; AB043997; BAB18161.1; -.  
 DR EMBL; AJ308378; CAC88115.1; -.  
 DR EMBL; AJ308379; CAC88115.1; JOINED.  
 DR EMBL; AJ308380; CAC88115.1; JOINED.  
 DR EMBL; AJ308381; CAC88115.1; JOINED.  
 DR EMBL; AJ308382; CAC88115.1; JOINED.  
 DR EMBL; AJ308383; CAC88115.1; JOINED.  
 DR EMBL; AJ308384; CAC88115.1; JOINED.  
 DR PIR; JC7502; JC7502.  
 DR Genew; HGNC:14025; SLC5A7.  
 DR GO; GO:0005624; C:membrane fraction; NAS.

DR GO; GO:0015220; F:choline transporter activity; TAS.  
 DR GO; GO:0008292; P:acetylcholine biosynthesis; NAS.  
 DR InterPro; IPR001734; Na/solut\_sympor.  
 DR Pfam; PF00474; SSF; 1.  
 DR PROSITE; PS50283; NA\_SOLUT\_SYMP\_3; 1.  
 SQ SEQUENCE 580 AA; 63203 MW; 66CB35496CB6E2D6 CRC64;

Query Match 100.0%; Score 2972; DB 4; Length 580;  
 Best Local Similarity 100.0%; Pred. No. 5.3e-205;  
 Matches 580; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY      1 MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTA 60
      |||
Db      1 MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTA 60

QY     61 TWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ 120
      |||
Db     61 TWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ 120

QY    121 IYGKRMGGLLFIPALMGEMFWAAAI FSALGATISVIIDVDMHISVIISALIATLYTLVGG 180
      |||
Db    121 IYGKRMGGLLFIPALMGEMFWAAAI FSALGATISVIIDVDMHISVIISALIATLYTLVGG 180

QY    181 LYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSDSEVYSW 240
      |||
Db    181 LYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSDSEVYSW 240

QY    241 LDSFLLMLGGIPWQAYFQRLSSSSATYAQVLSFLAAFGLVMAIPAILIGAIGASTDW 300
      |||
Db    241 LDSFLLMLGGIPWQAYFQRLSSSSATYAQVLSFLAAFGLVMAIPAILIGAIGASTDW 300

QY    301 NQTAYGLPDPKTTEEADMILPIVLQYLCVPYISFFGLGAVSAAVMSSADSSILSASSMFA 360
      |||
Db    301 NQTAYGLPDPKTTEEADMILPIVLQYLCVPYISFFGLGAVSAAVMSSADSSILSASSMFA 360

QY    361 RNIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQ 420
      |||
Db    361 RNIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQ 420

QY    421 LLCVLFVKGTNTYGAVAGYVSGFLRITGGEPYLYLQPLIFYPGYYPDDNGIYNQKFPFK 480
      |||
Db    421 LLCVLFVKGTNTYGAVAGYVSGFLRITGGEPYLYLQPLIFYPGYYPDDNGIYNQKFPFK 480

QY    481 TLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDVAVVARHSEENMDKTILVKKENIKLD 540
      |||
Db    481 TLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDVAVVARHSEENMDKTILVKKENIKLD 540

QY    541 ELALVKPRQSM T L S S T F T N K E A F L D V D S S P E G S G T E D N L Q 580
      |||
Db    541 ELALVKPRQSM T L S S T F T N K E A F L D V D S S P E G S G T E D N L Q 580

```

RESULT 2

Q9JMD7

ID Q9JMD7 PRELIMINARY; PRT; 580 AA.

AC Q9JMD7;

DT 01-OCT-2000 (TrEMBLrel. 15, Created)







Qy 121 IYGKRMGGLLFIPALMGEMFWAAAFSALGATISVIIDVDMHISVIISALIATLYTLVGG 180  
 |||:|||||  
 Db 121 IYGKRMGGLLFIPALMGEMFWAAAFSALGATISVIIDVDVNISVIVSALIATLYTLVGG 180  
 |||:|||||  
 Qy 181 LYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSSEVYSW 240  
 |||:|||||  
 Db 181 LYSVAYTDVVQLFCIFVGLWISVPFALSHPAVTDIGFTAVHAKYQSPWLGTIESVEVYTW 240  
 |||:|||||  
 Qy 241 LDSFLLMLGGIPWQAYFQRVLSSSSATYAQVLSFLAAFGLVMAIPAILIGAIGASTDW 300  
 ||:|||||  
 Db 241 LDNFFLLMLGGIPWQAYFQRVLSSSSATYAQVLSFLAAFGLVMAIPAILIGAIGASTDW 300  
 ||:|||||  
 Qy 301 NQTAYGLPDPKTTEEADMILPIVLQYLCVPYISFFGLGAVSAAVMSSADSSILSASSMFA 360  
 |||||  
 Db 301 NQTAYGYDPDKTKEEADMILPIVLQYLCVPYISFFGLGAVSAAVMSSADSSILSASSMFA 360  
 |||||  
 Qy 361 RNIYQLSFRQNASDKEIVWVMRITVVFVGASATAMALLTKTVYGLWYLSDDLVIYIIFPQ 420  
 |||:|||||  
 Db 361 RNIYQLSFRQNASDKEIVWVMRITVLFVGASATAMALLTKTVYGLWYLSDDLVIYIIFPQ 420  
 |||:|||||  
 Qy 421 LLCVLFVKGTNTYGAVAGYVSGFLRLITGGEPYLYLQPLIFYPGYYPDDNGIYNQKFPFK 480  
 |||||:|||||  
 Db 421 LLCVLFVKGTNTYGAVAGYVSGFLRLITGGEPYLYLQPLIFYPGYYSKNGIYNQKFPFK 480  
 |||||:|||||  
 Qy 481 TLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDVAVVARHSEENMDKTILVKENIKLD 540  
 ||:|||||  
 Db 481 TLMVTSFFTNICISYLAKYLFESGTLPPKLDVFDVAVVARHSEENMDKTILVRNENIKLN 540  
 ||:|||||  
 Qy 541 ELALVKPRQSMSTLSSTFTNKEAFLDVDSSPEGSGTEDNLQ 580  
 |||:|||||  
 Db 541 ELAPVKPRQSLTSLSTFTNKEALDVDSSPEGSGTEDNLQ 580  
 |||:|||||

RESULT 4

Q99PK3

ID Q99PK3 PRELIMINARY; PRT; 580 AA.  
 AC Q99PK3;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Sodium and chloride-dependent high-affinity choline transporter.  
 GN SLC5A7.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Apparsundaram S., Ferguson S.M., Blakely R.D.;  
 RT "Molecular cloning and characterization of human and murine high-  
 RT affinity choline transporters."  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AF276872; AAG36945.2; -.  
 DR MGD; MGI:1927126; Slc5a7.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0005215; F:transporter activity; IEA.  
 DR GO; GO:0006810; P:transport; IEA.

DR InterPro; IPR001734; Na/solut\_sympor.  
DR Pfam; PF00474; SSF; 1.  
DR PROSITE; PS50283; NA\_SOLUT\_SYMP\_3; 1.  
SQ SEQUENCE 580 AA; 63383 MW; DDBF58ED428270AF CRC64;

Query Match 94.0%; Score 2795; DB 11; Length 580;  
Best Local Similarity 92.6%; Pred. No. 2.7e-192;  
Matches 537; Conservative 23; Mismatches 20; Indels 0; Gaps 0;

```
Qy      1 MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTA 60
        | | | | | | : | | | : | | | | | | | | | | | | | | | | | | | | | | | |
Db      1 MPFHVEGLVAIILFYLLIFLVGIWAAWKTKNNGNPEERSEAIIVGGRDIGLLVGGFTMTA 60

Qy     61 TWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ 120
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     61 TWVGGGYINGTAEAVYVPGYGLAWAHAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFKQ 120

Qy    121 IYGKRMGGLLFIPALMGEMFWAAAFSALGATISVIIDVDMHISVIISALIAITLYTLVGG 180
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    121 IYGKRMGGLLFIPALMGEMFWAAAFSALGATISVIIDVDVNISVIVSALIAITLYTLVGG 180

Qy    181 LYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSSEVYSW 240
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    181 LYSVAYTDVVQLFCIFIGLWISVPFALSHPAVTDIGFTAVHAKYQSPWLGTIESVEVYTW 240

Qy    241 LDSFLLMLLGGIPWQAYFQRLSSSSATYAQVLSFLAAFGLVMAIPAILIGAIGASTDW 300
        | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    241 LDNFFLLMLLGGIPWQAYFQRLSSSSATYAQVLSFLAAFGLVMALPAICIGAIGASTDW 300

Qy    301 NQTAYGLPDPKKTTEEADMILPIVLQYLCVPYISFFGLGAVSAAVMSSADSSILSASSMFA 360
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    301 NQTAYGYDPDKTKKEADMILPIVLQYLCVPYISFFGLGAVSAAVMSSADSSILSASSMFA 360

Qy    361 RNIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIIVFPQ 420
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    361 RNIYQLSFRQNASDKEIVWVMRITVLVFGASATAMALLTKTVYGLWYLSSDLVYIIVFPQ 420

Qy    421 LLCVLFFVKGNTNTYGAVAGYVSGFLRITGGEPYLYLQPLIFYPGYPPDDNGIYNQKFPFK 480
        | | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    421 LLCVLFIKGTNTYGAVAGYIFGLFLRITGGEPYLYLQPLIFYPGYSDKNGIYNQRFPPFK 480

Qy    481 TLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDVAVVARHSEENMDKTILVKNENIKLD 540
        | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    481 TLSMVTSSFTNICVSYLAKYLFESGTLPPKLDVFDVAVVARHSEENMDKTILVRNENIKLN 540

Qy    541 ELALVKPRQSMTLSSTFTNKEAFLDVDSSPEGSGTEDNLQ 580
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    541 ELAPVKPRQSITLSSTFTNKEALLDVDSSPEGSGTEDNLQ 580
```

RESULT 5

Q9ESW5

ID Q9ESW5 PRELIMINARY; PRT; 580 AA.  
AC Q9ESW5;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

DE High affinity choline transporter.  
 GN SLC5A7 OR CHT1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BALB/cJ; TISSUE=Brain stem;  
 RA Bruess M.;  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BALB/cJ; TISSUE=Brain stem;  
 RA Wieland A., Bonisch H., Bruss M.;  
 RT "Molecular cloning of the human and murine high affinity choline  
 RT transporters and characterization of the human gene-structure."  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AJ401467; CAC03719.1; -.  
 DR MGD; MGI:1927126; Slc5a7.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0005215; F:transporter activity; IEA.  
 DR GO; GO:0006810; P:transport; IEA.  
 DR InterPro; IPR001734; Na/solut\_sympo.  
 DR Pfam; PF00474; SSF; 1.  
 DR PROSITE; PS50283; NA\_SOLUT\_SYMP\_3; 1.  
 SQ SEQUENCE 580 AA; 63331 MW; A4F1387CAA9EAAFE CRC64;

Query Match 93.9%; Score 2791; DB 11; Length 580;  
 Best Local Similarity 92.4%; Pred. No. 5.1e-192;  
 Matches 536; Conservative 24; Mismatches 20; Indels 0; Gaps 0;

Qy	1	MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTA	60
		:     :    :           :    :	
Db	1	MSFHVEGLVAIILFYLLIFLVGIWAAWTKNSGNPEEHSEAIIVGGRDIGLLVGGFTMTA	60
Qy	61	TWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ	120
Db	61	TWVGGGYINGTAVAVYGPGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ	120
Qy	121	IYGKRMGGLLFIPALMGEMFWAAAFSALGATISVIIDVDMHISVIISALIATLYTLVGG	180
Db	121	IYGKRMGGLLFIPALMGEMFWAAAFSALGATISVIIDVDVNISVIVSALIAILYTLVGG	180
Qy	181	LYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSSEVYSW	240
Db	181	LYSVAYTDVVQLFCIFIGLWISVPFALSHPAVTDIGFTAVHAKYQSPWLGTIESVEVYTW	240
Qy	241	LDSFLLMLGGIPWQAYFQRLSSSSATYAQVLSFLAAFGLVMAIPAILIGAIGASTDW	300
		:	
Db	241	LDNFLLMLGGIPWQAYFQRLSSSSATYAQVLSYLAAGGLVMALPAICIGAIGASTDW	300
Qy	301	NQTAYGLPDKTTEEADMILPIVLQYLCVPYISFFGLGAVSAAVMSSADSSILSASSMFA	360
Db	301	NQTAYGYDPDKTKEEADMILPIVLQYLCVPYISFFGLGAVSAAVMSSADSSILSASSMFA	360
Qy	361	RNIYQLSFRQNASDKEIVVMRITVFVFGASATAMALLTKTVYGLWYLSDDLVIYIVFPQ	420

Db	361	RNIYQLSFRQNASDKEIVVWMRITVLVFGASATAMALLTKTVYGLWYLSSDLVYIIIFPQ	420
Qy	421	LLCVLFVKGTNTYGAVAGYVSGFLFRITGGEPYLYLQPLIFYPGYPPDDNGIYNQKFPPK	480
Db	421	LLCVLFIKGTNTYGAVAGYIFGLFLRITGGEPYLYLQPLIFYPGYSDKNGIYNQRFPPK	480
Qy	481	TLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDVAVVARHSEENMDKTILVKENENIKLD	540
Db	481	TLSMVTSFFTNICVSYLAKYLFESGTLPPKLDVFDVAVVARHSEENMDKTILVRNENIKLN	540
Qy	541	ELALVKPRQSMTLSSTFTNKEAFLDVDSSPEGSGTEDNLQ	580
Db	541	ELAPVKPROSLTLSSTFTNKEALLDVDSSPEGSGTEDNLQ	580

08UWF0

```

ID      Q8UWF0      PRELIMINARY;      PRT;      584 AA.
AC      Q8UWF0;
DT      01-MAR-2002 (TrEMBLrel. 20, Created)
DT      01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT      01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE      High affinity choline transporter.
GN      CHT1.
OS      Torpedo marmorata (Marbled electric ray).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC      Elasmobranchii; Squala; Hypnosqualea; Pristiorajea; Batoidea;
OC      Torpediniformes; Torpedinoidei; Torpedinidae; Torpedo.
OX      NCBI_TaxID=7788;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Electric lobe;
RA      Guernonprez L., O'Regan S., Meunier F.M., Morot-Gaudry-Talarmain Y.;
RT      "Cyclosporin, FK506 and rapamycin inhibit neuronal choline uptake via
RT      calcineurin-dependent and independent mechanisms.";
RL      Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.
DR      EMBL; AJ420808; CAD12727.1; -.
DR      GO; GO:0016020; C:membrane; IEA.
DR      GO; GO:0005215; F:transporter activity; IEA.
DR      GO; GO:0006810; P:transport; IEA.
DR      InterPro; IPR001734; Na/solut_sympor.
DR      Pfam; PF00474; SSF; 1.
DR      PROSITE; PS50283; NA_SOLUT_SYMP_3; 1.
SQ      SEQUENCE      584 AA; 63660 MW; 995F937B01195A3D CRC64;

```

Qy	1	MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSG--SAEERSEAIIVGGRDIGLLVGGMFTM	58
		:: ::  ::       :    : :   :    :	
Db	1	MTVHIDGIVAIVLFYLLILFVGLWAAWKSKNTSMEGAMDRSEAIMIGGRDIGLLVGGMFTM	60
Qy	59	TATWVGGSYINGTAEAVYPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGVTMLDPF	118
Db	61	TATWVGGSYINGTAEAVYPGYGLAWAQPFGYALSLVIGGLFFAKPMRSGYVTMLDPF	120

Qy 119 QQIYGKRMGGLLFIPALMGEMFWAAAFSALGATISVIIDVDMHISVIISALIATLYTLV 178  
 ||:|||||:||||:||||:||||:||||:||||:||||:||||:||||  
 Db 121 QQMYGKRMGGLLFIPALLGEIFWSAAILSALGATLSVIVDININVSVVVSAVIAVLYTLV 180  
 Qy 179 GGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSEVY 238  
 |||||:||||:||||:||||:||||:||||:||||:||||:||||  
 Db 181 GGLYSVAYTDVVQLFCIFLGLWISIPFALLNPAVTDIIVTANQEVYQEPWVGNIQSKDSL 240  
 Qy 239 SWLDSFLLMLGGIPWQAYFQRVLSSSSSATYAQVLSFLAAGCLVMAIPAILIGAIGAST 298  
 |:|:|||||:||||:||||:||||:||||:||||:||||:||||:||||  
 Db 241 IWIDNFLLMLGGIPWQVYFQRVLSASSATYAQVLSFLAAGCVLMAIPSVLIGAIGTST 300  
 Qy 299 DWNQTAYGLPDPKTTTEADMILPIVLQYLCVPYISFFGLGAVSAAVMSSADSSILSASSM 358  
 |||||:||||:| | |||||:||||:|||||:|||||:|||||:|||||  
 Db 301 DWNQTSYGLPGPIGKNETDMILPIVLQHLCPYISFFGLGAVSAAVMSSADSSILSASSM 360  
 Qy 359 FARNIYQLSFRQNASDKEIVVWMRITVVFVGASATAMALLTKTVYGLWYLSDDLVIYIVIF 418  
 ||||| |:||| |||||:||||:||||:||||:||||:||||:||||:||||:||||  
 Db 361 FARNIYHLAFRQEASDKEIVVWMRITIFLFGGAATSMALLAQSIYGLWYLSDDLVIYIIF 420  
 Qy 419 PQLLCVLFVKGTNTYGAAGYVSGFLRITGGEPYLYLQPLIFYPGYYPD---DNGIYN 474  
 |||: |||||:||||:||||:| |||:|||||:||||:||||:| | |:|  
 Db 421 PQLISVLFVKGTNTYGSAGYIIGFLLRISGGEPYLYLQPLIFYPGYYPD---DNGIYN 480  
 Qy 475 QKFPFKTLAMVTSFLTNICISYLAHYLFESGTLPPKLDVFDVAVARHSEENMDKTILVKN 534  
 |:|||||:||||: ||| | :||| |||| | ||||| |:|:|:|:| |||| |:  
 Db 481 QRFPFKTMAMLFSLGNTGVSYLVKYLEVSGILPPKLDVFDVAVARHSEENMDKTILVKN 540  
 Qy 535 ENIKLDELALVKPRQSMSTLSSTFTNKEAFLDVDSSPE 571  
 :|| | || | | | :|:|: ||||| |:|: |  
 Db 541 DNITLSELVHVNPISASVSAAALTNKEAFEDIEPNPE 577

RESULT 7

Q8AV27

ID Q8AV27 PRELIMINARY; PRT; 377 AA.  
 AC Q8AV27;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Putative high affinity choline transporter 1 (Fragment).  
 GN CHT1.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Ciliary ganglion;  
 RX MEDLINE=22308883; PubMed=12421710;  
 RA Mueller F., Rohrer H.;  
 RT "Molecular control of ciliary neuron development: BMPs and downstream  
 RT transcriptional control in the parasympathetic lineage.";  
 RL Development 129:5707-5717(2002).  
 DR EMBL; AJ511267; CAD53475.1; -.

Query Match 56.5%; Score 1679; DB 13; Length 377;  
Best Local Similarity 85.4%; Pred. No. 1.9e-112;  
Matches 322; Conservative 28; Mismatches 27; Indels 0; Gaps 0;

Qy	144	AIFSALGATISVIIDVDMHISVIISALIATLYTLVGGGLYSVAYTDVVQLFCIFVGLWISV	203
Db	1	AIFSALGATISVITDINVNLSVIISALIATLYTLVGGGLYSVAYTDVVQLFCIFLGLWISV	60
Qy	204	PFALSHPAVADIGFTAVHAKYQKPWLGTVDSSSEVYSWLDSFLLMLGGIPWQAYFQRVLS	263
Db	61	PFALSNPAVTDIGFTAVHEVHQAPWLGTIGSLNIYTWLDNFFLLTFGGIPWQAYFQRVLS	120
Qy	264	SSSATYAQVLSFLAAFGCLVMAIPAILIGAIGASTDWNQTAYGLPDPKTTEEADMILPIV	323
Db	121	SSSATYAQVLSFLAAFGCIVMAIPAVLIGAIGASTAWNQTEYGVDPPIANKEADMILPIV	180
Qy	324	LQYLCPVYISFFGLGAVSAAMVSSADSSILSASSMFARNIYQLSFRQNASDKEIVVWMRI	383
Db	181	LQYLCPVYISFFGLGAVSAAMVSSADSSILSASSMFARNIYQLSFRQNASDREIVVWMRI	240
Qy	384	TVFVFGASATAMALLTKTVYGLWYLSSDLVYIIVIFPQLLCVLFVKGTNTYGAVAGYVSGL	443
Db	241	TVFLFGASATAMALLASSVYGLWYLSSDLVYIIIFPQLLCVLFIKGTNTYGAIAGYLFGL	300
Qy	444	FLRITGGEPYLYLQPLIFYPGYPPDDNGIYNQKFPFKTLAMVTSFLTNICISYLAKYLFE	503
Db	301	VLRITGGEPYLYLQPLIYYPGCYPDENNIYVQRFPFKTLAMLTSEFTNIIVSYLAKYLFG	360
Qy	504	SGTLPPKLDVFDVAVVAR	520
Db	361	SGTLPPKLDVFDVAVVAR	377

OX NCBI TaxID=7227;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkeley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
 RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,  
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,  
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,  
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,  
 RA Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,  
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
 RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,  
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
 RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,  
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;

```

RT "Sequencing of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanenvong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AE003723; AAF55583.2; -.
DR EMBL; AY047521; AAK77253.1; -.
DR FlyBase; FBgn0038641; CG7708.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR001734; Na/solut_sympoort.
DR Pfam; PF00474; SSF; 1.
DR PROSITE; PS50283; NA_SOLUT_SYMP_3; 1.
SO SEQUENCE 614 AA; 66893 MW; 71A77E1216360042 CRC64;

```



QY 184 VAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSSEVYSWLD 243  
 |||||:|||||:||||: ||| |: | : : : |:| |: : : :|  
 Db 178 VAYTDVIQLFCIFIGLWMCIPFAWSNEHVGS L-----SDLEVDWIGHVEPKKHLYIDY 231

QY 244 FLLLMLGGIPWQAYFQRLSSSSATYAQVLSFLAAGCLVMAIPAILIGAIGASTDWNQT 303  
 |||: ||||| ||||| ||||| :| ||:|||| ||:|||| :|||| :| ||:|  
 Db 232 GLLLVFVGGIPWQVYFQRLSSSKTAGRAQLLSYVAAAGCILMAIPPVLIGAIKATPWN 291

QY 304 AYGLPDPKTTTEADMILPIVLQYLCFVYISFFGLGAVSAAMSSADSSILSASSMFARNI 363  
 | | | | :| ||||:|||| | :||||||| |||||:||||:|||||:  
 Db 292 DYKGPYPLTVDETSMLPMVLQYLTDFVSFFGLGAVSAAMSSADSSVLSAASMFARNV 351

QY 364 YQLSFRQNASDKIIVVMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQLLC 423  
 |:| ||| ||: ||:||||: : | | || ||| :||| : |||||:| |||||  
 Db 352 YKLIFRQKASEMEIIVVMRVAIIVVGILATIMALTIPSIYGLWSMCSDLVYVILFPQLLM 411

QY 424 VL-FVKGTNTYGAVAGYVSGFLFRITGGEPYLYLQPLIFYPGYPPDDNGIYNQKFPFKTL 482  
 |:| | ||||:|: |:|:|||| | | ||| ||| | | |||:|:  
 Db 412 VVHFKKHCNTYGSLSAYIVALAIRLSGGEAILGLAPLIKYPGY---DEETKEQMFPERTM 468

QY 483 AMVTSFLTNICISYLAKYLFESGTLPPKLDVFDVAVV 518  
 ||: | :| | :|: | :|||| ||| | | ||  
 Db 469 AMLLSLVTLISVSWWTMMFESGKLPPSYDYFRCVV 504

# RESULT 9

## Q9GPB1

ID Q9GPB1 PRELIMINARY; PRT; 579 AA.  
 AC Q9GPB1;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Choline cotransporter.  
 OS Limulus polyphemus (Atlantic horseshoe crab).  
 OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Merostomata; Xiphosura;  
 OC Limulidae; Limulus.  
 OX NCBI\_TaxID=6850;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21261948; PubMed=11368908;  
 RA Wang Y., Cao Z., Newkirk R.F., Ivy M.T., Townsel J.G.;  
 RT "Molecular cloning of a cDNA for a putative choline co-transporter  
 RT from Limulus CNS.";  
 RL Gene 268:123-131(2001).  
 DR EMBL; AY011119; AAG41055.1; -.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0005215; F:transporter activity; IEA.  
 DR GO; GO:0006810; P:transport; IEA.  
 DR InterPro; IPR001734; Na/solut\_sympor.  
 DR Pfam; PF00474; SSF; 1.  
 DR PROSITE; PS50283; NA\_SOLUT\_SYMP\_3; 1.  
 SQ SEQUENCE 579 AA; 62937 MW; FE7F29D4FAF47F04 CRC64;

Query Match 51.5%; Score 1530.5; DB 5; Length 579;  
 Best Local Similarity 52.0%; Pred. No. 1.5e-101;  
 Matches 305; Conservative 115; Mismatches 134; Indels 33; Gaps 11;





```

      :||| ||::| | :||:|:: ||| : | || ||| :::||||| :|||:
Db      358 ASMFANHNIWKLTI RPHASEKEV IIVMRIA IICV GIMATIMALT IQSIYGLWYLCADLVYV 417
Qy      416 VIFPQLLCVLFVKGTNTY GAVAGYVSGLFLRITGGEPYLYLQPLIFYPGYPPDDNGIYNQ 475
      ::||| ||||::: :|||::||| || ||: |||| : | || | :|: |
Db      418 ILFPQLLCVVMPRSNTY GSLAGYAVGLVLR LIGGEPLVSLPAFFHYPMY---TDGV--Q 472
Qy      476 KFPEKTLAMVTSFLT NICISYLAKYLFESGTLPPKLDVFDVAVV---ARHSEENMDKTILV 532
      |||:| ||::| | :| :: ||:| | |: || || | | : |
Db      473 YFPFRTTAMLSSMATIYIVSIQSEKLFKSGRLSPEWDVMGCVVNIPIDHVP LPSDVSEFAV 532
Qy      533 KNENIKL-----DELALVKPRQSMTLSSTFTN 559
      :| : : || | : | : || :|
Db      533 SSETLNMKAPNGTPAPVHPNQPSDENTLLHPYSDQSYYSTNSN 576

```

# RESULT 11

## Q7UFM6

```

ID      Q7UFM6          PRELIMINARY;      PRT;      484 AA.
AC      Q7UFM6;
DT      01-OCT-2003 (TrEMBLrel. 25, Created)
DT      01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE      High affinity choline transporter.
GN      CHT1 OR RB8472.
OS      Rhodopirellula baltica.
OC      Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;
OC      Planctomycetaceae; Pirellula.
OX      NCBI_TaxID=117;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=1;
RX      MEDLINE=22735913; PubMed=12835416;
RA      Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,
RA      Ludwig W., Gade D., Beck A., Borzym K., Heitmann K., Rabus R.,
RA      Schlesner H., Amann R., Reinhardt R.;
RT      "Complete genome sequence of the marine planctomycete Pirellula sp.
RT      strain 1.";
RL      Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303(2003).
DR      EMBL; BX294147; CAD78656.1; -.
KW      Complete proteome.
SQ      SEQUENCE      484 AA;      52674 MW;      79AB0135F18FEBB2 CRC64;

```

```

Query Match          14.2%;      Score 422.5;      DB 16;      Length 484;
Best Local Similarity 27.6%;      Pred. No. 3.9e-22;
Matches 141;      Conservative 95;      Mismatches 220;      Indels 55;      Gaps 14;

```

```

Qy      7 GLIAIIVFYLLI-LLVGIWAAWRTKN SSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGG 65
      ||| :: ||| : :|: || | | :: :| || : | : : ||| |
Db      2 GLIAAVLAYLLLTIAIGLLAARRVGN-----AQDFMVAGRS LPLYMNFACVFATWFG- 53
Qy      66 GYINGTAEAVY-----VPGYGL-AWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQ 119
      || | | || | | |:|: |:| |||: :|: | ::
Db      54 -----AETVLSVSATFAGQGLRAIPGDPFGFSICLV LVALFFARAFYRMDLLTIGDFYR 107
Qy      120 QIYGKRMGGLLFIPALMGEMFWAAAI FSALGATISVI-----IDVDMHISVIISALIAT 173
      : ||: : | : : ||| :||| |||: : :: ::| |

```

Db	108	KRYGRSIEVLTSVVISASYLGWAAAQLTALGLVISVLGKGIGYETLTINNGIVIGFTIVA	167
Qy	174	LYTLVGGLYSVAYTDVVQLFCIFVGLW-ISVPFALSHPAVADIGFTAVHAKYQKPWLGTV	232
		::  ::    :::     ::  :     :  :: :  : : :	
Db	168	FYTVMGGMWSVALTDMIQTFVIIIGLLVSVYMAHAAGGVSVVIESARESGLQVFPDWG	227
Qy	233	DSSEVYSWLDSFLLMLGGIPWQAYFQRVLSSSSATYAQVLSFLAA-FGCLVMAIPA-IL	290
		: : ::                  :   :      : :	
Db	228	QSGQWWIYIGGFLTAALGSIQQQDVVFQRVTSAKDERTAMTGTLLGGMFYCMFAFVPMFIA	287
Qy	291	IGAIGASTDWNQTAYGLPDPKTTTEADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADS	350
		:     :      :   ::     : :: :	
Db	288	YAAVVIDPDHLQQF----NSDDLREVQRTLPHAVIQSTPFWWQTVFLGALVSAILSTASG	343
Qy	351	SILSASSMFARNIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLT-KTVYGLWYLS	409
		:: :   :  : :    :    :: :   : :        :  :  : :	
Db	344	TLLAPSSLIVENVIR-PFRSDLDKNNMLRWLRIVLLMF GALALHQA LTSNNTMYEMIQQA	402
Qy	410	SDLVYIVIFPQLLCVLFVKGTNTYGAVAGYVSGFLFRITGGEPYLYLQPLIFYPGYYPDD	469
		: :    :       :     : : :  :	
Db	403	YSVPLVGALVPLAVGLYWKRA TRGAMASIVSGVATWLA----FEYMLPEFLIPS-----	453
Qy	470	NGIYNQKFPFKTLAMVTSFLTNICISYLA KY	500
		: : :     : :     :	
Db	454	-----OLMGLAASFLAMVVVSLLDKF	474

```

RESULT 12
Q8EXG7
ID   Q8EXG7          PRELIMINARY;          PRT;   462 AA.
AC   Q8EXG7;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Probable sodium:solute symporter.
GN   LB245.
OS   Leptospira interrogans.
OC   Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.
OX   NCBI_TaxID=173;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=56601 / Serogroup Icterohaemorrhagiae / Serovar lai;
RA   Ren S.;
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE011612; AAN51804.1; -.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0005215; F:transporter activity; IEA.
DR   GO; GO:0006810; P:transport; IEA.
DR   InterPro; IPR001734; Na/solut_symport.
DR   Pfam; PF00474; SSF; 1.
DR   PROSITE; PS50283; NA_SOLUT_SYMP_3; 1.
KW   Complete proteome.
SQ   SEQUENCE   462 AA;  50487 MW;  C9B0104065514C68 CRC64;

Query Match          13.6%;  Score 405.5;  DB 16;  Length 462;
Best Local Similarity 27.5%;  Pred. No. 6.2e-21;
Matches 134;  Conservative 103;  Mismatches 204;  Indels 47;  Gaps 16

```

Qy 8 LIAI-IVFYLL-ILLVGIWAARWTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGG 65  
 Db 1 MLAISVIFYLFTTILIGAVASRFVSD-----SKDYVLAGRRLPLFLASSALFATWFGS 53

Qy 66 GYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQIYGKR 125  
 Db 54 ETLLG-ASSRFVEDGILGVIEDPFGAALCLFLVGLFFARPLYRMNILTFGDFYKNRFGRR 112

Qy 126 ---MGGLLFIPALMGEMFWAAAFSALGATISVIIDVDMHISVIISALIATLYTLVGGLY 182  
 Db 113 AEILSSVFMIPSYFG---WIAAQFVALGIIFHSLADIPVSTGIIAGAGVVLIIYTVTGGMW 169

Qy 183 SVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKP----WLGTVDSSEVY 238  
 Db 170 AISLTDFLQTVLIVLGLSYLV-WDLSSKAG---GIEKILAS-TKPGFFRFFPEMNAKSIF 224

Qy 239 SWLDSFLLMLGGIPWQAYFQVLSSSSATYAQVLSFLAAGCLVMA-IPAILIGAIGAS 297  
 Db 225 AYIAAWMTIGLGSIPQQDIFQRMASKSEKVAVYSSLLGSFFYLSVAFLP--LIAVLCAR 282

Qy 298 TDWNQTAYGLPDPKTTEEADMILPIVLQYLCVPYISFFGLGAVSAAVMSSADSSILSASS 357  
 Db 283 KIYPEIA-----KEDAQMILPKTVLTHTGFLTQILFFGALLSAVMSTASGAILASAS 334

Qy 358 MFARNIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIVI 417  
 Db 335 VLGENVIRPFFKK-TSERTLLRLFRLSVIAITLVSLSMANTKSNIYELVSQASALSLVSL 393

Qy 418 FPQLLCVLFVKGTNTYGAAGYVSG-----LFLRITGGEPYLYLQPLIFYPGYYPD 468  
 Db 394 FIPLVAGLFRKNSTSTGAIFSMIVGFCTWFLCNILSLEIPASIPGLISSWIALYLGDWME 453

Qy 469 DNGIYNQK 476  
 Db 454 HRG-YIQK 460

# RESULT 13

Q8Y273

ID Q8Y273 PRELIMINARY; PRT; 479 AA.  
 AC Q8Y273;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Probable sodium/solute symporter transmembrane protein.  
 GN RSC0463 OR RS04434.  
 OS Ralstonia solanacearum (Pseudomonas solanacearum).  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 OC Burkholderiaceae; Ralstonia.  
 OX NCBI\_TaxID=305;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GMI1000;  
 RX MEDLINE=21681879; PubMed=11823852;  
 RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,  
 RA Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,

RA Chandler M., Choisine N., Claudel-Renard C., Cunnac S., Demange N.,  
 RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,  
 RA Siguier P., Thebault P., Whalen M., Wincker P., Levy M.,  
 RA Weissenbach J., Boucher C.A.;  
 RT "Genome sequence of the plant pathogen *Ralstonia solanacearum*.";  
 RL Nature 415:497-502(2002).  
 DR EMBL; AL646059; CAD13991.1; -.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0005215; F:transporter activity; IEA.  
 DR GO; GO:0006810; P:transport; IEA.  
 DR InterPro; IPR001734; Na/solut\_sympoort.  
 DR Pfam; PF00474; SSF; 1.  
 DR PROSITE; PS50283; NA\_SOLUT\_SYMP\_3; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 479 AA; 52091 MW; 560962E411DBC9B8 CRC64;

Query Match 12.8%; Score 381.5; DB 16; Length 479;  
 Best Local Similarity 28.2%; Pred. No. 3.4e-19;  
 Matches 128; Conservative 85; Mismatches 188; Indels 53; Gaps 16;

Qy	11	IIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGYING	70
		:  ::: : :      :  :     :     :      :	
Db	6	VIVYWVISVGIGLWAALRVRNTAD-----FAVAGRGLPFYVVTATVFATWFGSETVLG	58
Qy	71	TAEAVYVPGYGLAWAQA-PIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQIYGKRMGGL	129
		::                      : :   : : :	
Db	59	-IPAVFLK-EGLHGVVADPFSSSLCLILVGLFFARPLYRMNLLTIGDFYRNRFRGVAEVL	116
Qy	130	LFIPALMGEMFWAAAFSALGATISVIID--VDMHISVIISALIATLYTLVGGLYSVAYT	187
		: : :           :   : :     :        :	
Db	117	TTLCIVVSYLGWVAAQIKALGLVFYTVSDGGLSQQTGMMIGAASVLVYTLFGGMWSVAVT	176
Qy	188	DVQVLFQIFVGLWISVPFALSHPAVADIGFTAVHA----KYQKPWLGTVDSSSEVYSWLDS	243
		:   :   : : :     :     : : :     : : :	
Db	177	DFIQMIIIVIGM-MYIGWEVSGQA-GGVATVVAHASAAGKFS--FWPAFNPIEVIGFVTA	232
Qy	244	FLLMLGGIPWQAYFQRVLSS-----SSATYAQVLSFLAAFGLVMAIPAILIGAIGA	296
		: : :                : :         : :	
Db	233	WITMMLGSIPQQDVFORVTSSRTERIAGTASVLGGVLYFLFAFIPMFLAYSATLI-----	287
Qy	297	STDWNQTAYGLPDPK-----TTEADMILP-IVLQYLCVPYISFFGLGAVSAAVMSSADS	350
		: : :    :    :   :   : :     : :	
Db	288	-----DPQMVARVINTDSQLILPKLVLEH-APLVAQVMFFGALLSAIKSCASA	334
Qy	351	SILSASSMFARNIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTK-TVYGLWYLS	409
		: :       : :     :   :           : : : :	
Db	335	TLLAPSVTFAENVLR-PMLPRMDDKRFLRVMQAVVLVFTALVTLFALNSHLSIFHMVENA	393
Qy	410	SDLVYIVIFPQLLCVLFVKGTNTYGAVAGYVSGL	443
		: :           :	
Db	394	YKVTLVAAAFVPLAFGLFWKRATRQGGLLAIALGL	427

RESULT 14

Q9V2P3

ID Q9V2P3 PRELIMINARY; PRT; 492 AA.

AC Q9V2P3;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Proline symporter (Proline permease).  
 GN PUTP-3 OR PYRAB00320 OR PAB2354.  
 OS Pyrococcus abyssi.  
 OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
 OC Pyrococcus.  
 OX NCBI\_TaxID=29292;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GE5 / Orsay;  
 RA Heilig R.;  
 RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome  
 structure and evolution.";  
 RL Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AJ248283; CAB48955.1; -.  
 DR PIR; D75188; D75188.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0005215; F:transporter activity; IEA.  
 DR GO; GO:0006810; P:transport; IEA.  
 DR InterPro; IPR001734; Na/solut\_symport.  
 DR Pfam; PF00474; SSF; 1.  
 DR TIGRFAMS; TIGR00813; sss; 1.  
 DR PROSITE; PS50283; NA\_SOLUT\_SYMP\_3; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 492 AA; 53457 MW; A7C72B1AF29282B3 CRC64;

Query Match 11.6%; Score 344; DB 17; Length 492;  
 Best Local Similarity 24.2%; Pred. No. 1.7e-16;  
 Matches 132; Conservative 99; Mismatches 196; Indels 118; Gaps 25;

Qy 8 LIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGY 67  
 |:| : :| :|| :| ||| : | ||| : : : ::  
 Db 14 LVAFLETLILPILVGFYAMKRTKS-----EEDFFVGGGRAMDKITVALSAVSSGRSSWL 66  
  
 Qy 68 INGTAEEAVYVPGYGLAWAQAPIGYSLS----LILGGLFFAKPMRSKGYVTMLDPFQQIYG 123  
 : | : | | | | :|| : : : | : | : | : :  
 Db 67 VLGLSGMAYKMGVTAVW--AAVGVIYAEMFQFVYMGIRLRKFSEFNAITVPDYFEARFR 124  
  
 Qy 124 K-----RMGG----LLFIPALMGEMFWAAAFSALGATISVIIDVDMHISVIISALIATL 174  
 |: : :| :| | | | |:| : : : :|| | : :  
 Db 125 DTSKILRIAASIIIIIFLTSYVGAQFNAGA-----KTLSTALGISIFTALMISVIMIV 178  
  
 Qy 175 YTLVGGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFT-----AVHAKYQK 226  
 | :|| :||| ||: : :|| : | ||| :| | | : |  
 Db 179 YMILGGFIAVAYNDVIRAVIMIIGLVV----LPVIAVAKVGGTEEVVLKVLHALDPKLIN 233  
  
 Qy 227 PW---LGTVDSSSEVYSWLDSFLLMLG-GIPWQAY-FQVLSSSSATYAQVLSFLAAFGC 281  
 || | | | :| | | | : | :| : : : |  
 Db 234 PWAFGAGVVIG-----FLGIGFGSPGQPHIIVRYMSIDDPNKL RVSTVVGTFWN 282  
  
 Qy 282 LVMAIPAILIGAIGASTDWNQTAYGLPDPKTT--EEADMILP-IVLQYLCVPYISFFGLG 338  
 :|:| || :| | : :|| : :|| | : || | : |  
 Db 283 VVLAWGAIFVGLAGRAI-----VPDVSQPLPGKNAEMIYPYLSAQYFPPILYGIL-IG 333  
  
 Qy 339 AVSAAVMSSADSSILSASSMFARNIYQLSFRQNA--SDKEIVWVMRITVFVFGASATAMA 396



```

      : ||::|:| | :| :| :::| : : :| :| | | :|
Db      334 GIFAAILSTADSQLLVASTVVKDLYQEVKKGKTKIDEKTALTISRVTVLVVGFLAAILA 393
QY      397 LLTKTVYGLWYLSSDLVY-IVIF-----PQLLCVLFVKGTNTYGAVAGYVSGFL 445
      |::|::| :| | :| :| | :| :| :|
Db      394 -----YVAKDIIFWVFLFAWGGLGASFGLTLILSLYWKGTTKWGVLAGMIVGTIT 443
QY      446 RITGGEPYLYLQPLIFYPGYYPDDNGIYNQKFPFKTLAMVTSFLTNICISYLAKYLFESG 505
      | | |::| :| :| :| :| :| :| :|
Db      444 TIVW---KLYLKPI-----TGLY-ELVP---AFIFSLIATIIIVSMITK----- 479
QY      506 TLPPK 510
      ||:
Db      480 --PPE 482

```

RESULT 15

Q81Y52

```

ID      Q81Y52          PRELIMINARY;          PRT;          492 AA.
AC      Q81Y52;
DT      01-JUN-2003 (TrEMBLrel. 24, Created)
DT      01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE      Sodium/proline symporter family protein.
GN      BA3705.
OS      Bacillus anthracis (strain Ames).
OC      Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX      NCBI_TaxID=198094;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=22608414; PubMed=12721629;
RA      Read T.D., Peterson S.N., Tourasse N., Baillie L.W., Paulsen I.T.,
RA      Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,
RA      Holtzapple E.K., Okstad O.A., Helgason E., Rilstone J., Wu M.,
RA      Kolonay J.F., Beanan M.J., Dodson R.J., Brinkac L.M., Gwinn M.,
RA      DeBoy R.T., Madpu R., Daugherty S.C., Durkin A.S., Haft D.H.,
RA      Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,
RA      Benton J.L., Mahamoud Y., Jiang L., Hance I.R., Weidman J.F.,
RA      Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C.,
RA      Hazen A., Cline R., Redmond C., Thwaite J.E., White O., Salzberg S.L.,
RA      Thomason B., Friedlander A.M., Koehler T.M., Hanna P.C., Kolsto A.-B.,
RA      Fraser C.M.;
RT      "The genome sequence of Bacillus anthracis Ames and comparison to
RT      closely related bacteria.";
RL      Nature 423:81-86(2003).
DR      EMBL; AE017035; AAP27454.1; -.
DR      TIGR; BA3705; -.
DR      GO; GO:0016020; C:membrane; IEA.
DR      GO; GO:0005215; F:transporter activity; IEA.
DR      GO; GO:0006810; P:transport; IEA.
DR      InterPro; IPR001734; Na/solut_symport.
DR      Pfam; PF00474; SSF; 1.
DR      TIGRFAMs; TIGR00813; sss; 1.
DR      PROSITE; PS50283; NA_SOLUT_SYMP_3; 1.
KW      Complete proteome.
SQ      SEQUENCE 492 AA; 53891 MW; E2377D735C1A90F9 CRC64;

```

Query Match 11.2%; Score 334; DB 16; Length 492;  
 Best Local Similarity 22.9%; Pred. No. 9e-16;  
 Matches 129; Conservative 101; Mismatches 219; Indels 114; Gaps 17;

```

Qy      5 VEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVG 64
      :| :::: ::  :| :| |: :| :      ::||| :| |  :  |: :
Db      3 IEIMVSLAIYMAGMLYIGYWSYKKTSDLSD-----YMLGGRGLGPAVTALSAGASDMS 55

Qy     65 GGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGG----LFFAKPMR-----SKGYVTML 115
      | : |  |:| |  |      ::|::|  | | :|  :  ::|
Db     56 GWMLMGLPGAMYATGLSSVW-----IAIGLLIGAYANYLILAPRLRITYTEVANDSITIP 109

Qy    116 DPFQQIYGKRMGGLLFIPA---LMGEMFWAAAFSALGATISVIIDVDMHISVIISALIA 172
      | : :  |  | |: |  |: |::|  : |  ::| | : :: :
Db    110 DFLENRFKDRTKILRFVSAIVILVFFTFYASAGLVSGGRLFENSFNLDYKIGLFVTVGVV 169

Qy    173 TLYTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIG-----FTAVHAKYQKP 227
      ||| ||  :|::|| ||  :|: | : ||  | |:|  | :
Db    170 VAYTLFGGFLAVSWTDFVQGCIMFIAL-VLPPIV----AFTDVGGVTETFTNTIK----- 218

Qy    228 WLGTVDSSEVYSWLDSFLLMLGGIPW-QAYF-----QRLSSSSSATYAQVLSFLAAG 280
      ||:| :  :  : :| ::  : |  ||  | :: :|  :  :
Db    219 ---QVDASHLDMFKGTTILGIISFLAWGLGYFGQPPIIVRFMAITSIKDLKTSRRIGIGW 275

Qy    281 CLVMAIPAILIGAIGASTDWNQTAYGLPDKTTEEADMILPIVLQYLCVPYISFFGLGAV 340
      :  | |:| | :|  ||  :  | :: :|:  |  ||: | | |:
Db    276 MTISIIGAMLTGLVG-----IAYYAKNNATLQDPEMVFVTFSNILFHPYITGFLLSAI 328

Qy    341 SAAVMSSADSSILSASSMFARNIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTK 400
      |::||| | :| ||  : |: ||: |||||::|::| |  | :|
Db    329 LASIMSSISSQLLVISSAVTEDFYKTFRRKASDKELVFIGRLSVLVVAMIAVLA---- 384

Qy    401 TVYGLWYLSSDLVYIVI-----FPQLLCVLFVKGTNTYGAVAGYVSGFLFRITG 449
      |  || :  ::  | :|  |: | || :| :|| : |  : ||
Db    385 -----YHPSDTILTLVGYAWAGFGSAFGPAILLSLYWKRTNKGVLAGMIVGALVVITW 438

Qy    450 GE-PYLYLQPLIFYPGYPPDDNGIYNQKFFPKTLAMVTSFLTNICISYLAKYLFESGTL 508
      : | |  ||::  | | | :| :|
Db    439 VQIPSLKASMYEMVPGFF-----CSLLAVIIVSLVTK----- 470

Qy    509 PKLDVFDVAVVARHSEENMDKTIL 531
      : | | | | | : :|
Db    471 -----EPVKAIHREFNEMEAVL 487

```

Search completed: March 22, 2004, 15:35:00  
 Job time : 113 secs

OM protein - protein search, using sw model

Run on: March 22, 2004, 13:57:40 ; Search time 26 Seconds  
 (without alignments)  
 1161.565 Million cell updates/sec

Title: US-10-069-541-6  
 Perfect score: 2972  
 Sequence: 1 MAFHVEGLIAIIVFYLLILL.....EAFLDVDSSPEGSGTEDNLQ 580

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	308.5	10.4	662	1	SL51_RABIT	P11170 oryctolagus
2	308	10.4	670	1	SL52_RAT	P53792 rattus norv
3	306	10.3	664	1	SL51_HUMAN	P13866 homo sapien
4	306	10.3	665	1	SL51_RAT	P53790 rattus norv
5	303.5	10.2	660	1	SL54_PIG	P31636 sus scrofa
6	298	10.0	672	1	SL52_RABIT	P26430 oryctolagus
7	294	9.9	659	1	SL54_HUMAN	Q9ny91 homo sapien
8	294	9.9	664	1	SL51_SHEEP	P53791 ovis aries
9	293.5	9.9	543	1	SGLT_VIBPA	P96169 vibrio para
10	292	9.8	672	1	SL52_HUMAN	P31639 homo sapien
11	290	9.8	656	1	SL54_MOUSE	Q9et37 mus musculu
12	289	9.7	718	1	SL53_MOUSE	Q9jkz2 mus musculu
13	285	9.6	492	1	OPUE_BACSU	O06493 bacillus su
14	278.5	9.4	718	1	SL53_CANFA	P31637 canis famil
15	275	9.3	718	1	SL53_BOVIN	P53793 bos taurus
16	272.5	9.2	718	1	SL53_HUMAN	P53794 homo sapien
17	260.5	8.8	504	1	POTP_HAEIN	P45174 haemophilus

18	260.5	8.8	605	1	SL51_PIG	P26429	sus scrofa
19	253	8.5	502	1	PUTP_ECOLI	P07117	escherichia
20	249	8.4	502	1	PUTP_SALTY	P10502	salmonella
21	239.5	8.1	484	1	PANF_HAEIN	P44963	haemophilus
22	239	8.0	549	1	YJCG_ECOLI	P32705	escherichia
23	231	7.8	643	1	SL55_HUMAN	Q92911	homo sapien
24	229.5	7.7	636	1	SL56_RABIT	Q9xt77	oryctolagus
25	226	7.6	571	1	YIDK_ECOLI	P31448	escherichia
26	219.5	7.4	634	1	SL56_RAT	O70247	rattus norv
27	218	7.3	618	1	SL55_RAT	Q63008	rattus norv
28	209	7.0	635	1	SL56_HUMAN	Q9y289	homo sapien
29	206.5	6.9	483	1	PANF_ECOLI	P16256	escherichia
30	198.5	6.7	735	1	DUR3_YEAST	P33413	saccharomyc
31	198	6.7	516	1	YWCA_BACSU	P39599	bacillus su
32	137	4.6	657	1	NU5M_EMENI	P11628	emericella
33	132	4.4	482	1	ARCD_PSEAE	P18275	pseudomonas
34	132	4.4	492	1	YCLF_BACSU	P94408	bacillus su
35	132	4.4	499	1	MVIN_AQUAE	O67658	aquifex aeo
36	131	4.4	570	1	NU5M_PARTE	P15584	paramecium
37	125.5	4.2	499	1	NQOE_PARDE	P29926	paracoccus
38	125	4.2	476	1	YAAJ_ECOLI	P30143	escherichia
39	124.5	4.2	666	1	NU5M_CHOCR	P48920	chondrus cr
40	123	4.1	517	1	QOX1_SULAC	P98004	sulfolobus
41	121	4.1	692	1	NU5C_MARPO	P06264	marchantia
42	120	4.0	687	1	CSTA_HELPJ	Q9zk47	helicobacte
43	120	4.0	732	1	YLH3_SCHPO	Q9hgg3	schizosacch
44	120	4.0	741	1	YBIO_ECOLI	P75783	escherichia
45	119	4.0	468	1	MLEN_BACSU	P54571	bacillus su

#### ALIGNMENTS

##### RESULT 1

##### SL51\_RABIT

ID SL51\_RABIT STANDARD; PRT; 662 AA.  
 AC P11170;  
 DT 01-JUL-1989 (Rel. 11, Created)  
 DT 01-JUL-1989 (Rel. 11, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Sodium/glucose cotransporter 1 (Na+)/glucose cotransporter 1)  
 DE (High affinity sodium-glucose cotransporter).  
 GN SLC5A1 OR SGLT1.  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_TaxID=9986;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=New Zealand white;  
 RX MEDLINE=88065856; PubMed=2446136;  
 RA Hediger M.A., Coady M.J., Ikeda T.S., Wright E.M.;  
 RT "Expression cloning and cDNA sequencing of the Na+/glucose co-  
 RT transporter."  
 RL Nature 330:379-381(1987).  
 RN [2]  
 RP SEQUENCE FROM N.A.

RC STRAIN=New Zealand white; TISSUE=Kidney cortex;  
 RX MEDLINE=91223090; PubMed=2025641;  
 RA Morrison A.I., Panayotova-Heiermann M., Feigl G., Schoelermann B.,  
 RA Kinne R.K.H.;  
 RT "Sequence comparison of the sodium-D-glucose cotransport systems in  
 RT rabbit renal and intestinal epithelia.";  
 RL Biochim. Biophys. Acta 1089:121-123(1991).  
 CC -!- FUNCTION: Actively transports glucose into cells by Na(+) co-  
 CC transport with a Na(+) to glucose coupling ratio of 2:1.  
 CC -!- FUNCTION: Efficient substrate transport in mammalian kidney is  
 CC provided by the concerted action of a low affinity high capacity  
 CC and a high affinity low capacity Na(+)/glucose cotransporter  
 CC arranged in series along kidney proximal tubules.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -!- TISSUE SPECIFICITY: Found predominantly in intestine and in outer  
 CC renal medulla.  
 CC -!- DISEASE: Mutation of Asp-28 is implicated in glucose/galactose  
 CC malabsorption.  
 CC -!- SIMILARITY: BELONGS TO THE SODIUM:SOLUTE SYMPORTER FAMILY (SSF).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X06419; CAA29727.1; -.  
 DR EMBL; X55355; CAA39040.1; -.  
 DR PIR; S00515; A37226.  
 DR InterPro; IPR001734; Na/solut\_symport.  
 DR Pfam; PF00474; SSF; 1.  
 DR TIGRFAMS; TIGR00813; sss; 1.  
 DR PROSITE; PS00456; NA\_SOLUT\_SYMP\_1; 1.  
 DR PROSITE; PS00457; NA\_SOLUT\_SYMP\_2; 1.  
 DR PROSITE; PS50283; NA\_SOLUT\_SYMP\_3; 1.  
 KW Transport; Sugar transport; Transmembrane; Sodium transport; Symport;  
 KW Glycoprotein.  
 FT DOMAIN 1 28 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 29 47 POTENTIAL.  
 FT DOMAIN 48 64 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 65 85 POTENTIAL.  
 FT DOMAIN 86 105 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 106 126 POTENTIAL.  
 FT DOMAIN 127 171 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 172 191 POTENTIAL.  
 FT DOMAIN 192 208 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 209 229 POTENTIAL.  
 FT DOMAIN 230 270 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 271 291 POTENTIAL.  
 FT DOMAIN 292 314 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 315 334 POTENTIAL.  
 FT DOMAIN 335 423 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 424 443 POTENTIAL.  
 FT DOMAIN 444 455 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 456 476 POTENTIAL.

FT	DOMAIN	477	526	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	527	547	POTENTIAL.
FT	DOMAIN	548	640	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	641	661	POTENTIAL.
FT	DOMAIN	662	662	EXTRACELLULAR (POTENTIAL).
FT	CARBOHYD	248	248	N-LINKED (GLCNAC. . .).
FT	SITE	43	43	IMPLICATED IN SODIUM COUPLING
FT				(BY SIMILARITY).
FT	SITE	300	300	IMPLICATED IN SODIUM COUPLING
FT				(BY SIMILARITY).
SQ	SEQUENCE	662 AA;	73079 MW;	03F55A0309CBBE01 CRC64;

Query Match 10.4%; Score 308.5; DB 1; Length 662;  
 Best Local Similarity 23.4%; Pred. No. 2.6e-14;  
 Matches 154; Conservative 110; Mismatches 238; Indels 155; Gaps 26;

Qy	11	IIVFYLLILLVGIWAAWRTKNSGSAERSEAIIVGGRDIGLLVGGFTMTATWVGGGYING	70
		::: :::   :    :       : :     :   :   :   :	
Db	32	IVIYFLVVMVAVGLWAMFST-NRGTV---GGFFLAGRSMVWVPIGASLFASNIGSGHFVG	86
Qy	71	TAEAVYVPGYGLAWAQAPIGYS-----LSLILGGLFFAKPMRSKGYVTMLDPFQQIY-GK	124
		: : :     :   :       :   :	
Db	87	LA-----GTGAASGIATGGFEWNALIMVVVLGWVFPVIYIRA-GVVTMPEYLQKRFGGK	139
Qy	125	RMGGLLFIPALMGEMFW--AAAFSALGAT-ISVIIDVDMHISVIISALIATLYTLVGGL	181
		:     :   :   :           : : :   : :	
Db	140	RIQIYLSILSLLLYIFTKISADIFS--GAIFIQLTLGLDIYVAIIILLVITGLYTITGGL	197
Qy	182	YSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKY-----Q	225
		:       :   :         :	
Db	198	AAVIYTDTLQTAIMMVGSVILTGFAFHEVG---GYEAFTEKYMRAIPSQISYGNTSIPQ	253
Qy	226	KPWLGTVDSSSEVYSWLDSFLLMLGGIPW-----QAYFQRVLSSSSA	267
		:   :   :   :                 : :	
Db	254	KCYTPREDAFHI-----FRDAITGDIPWPGLVFGMSILTLWYWCTDQVIVQRCLSAKNL	307
Qy	268	TYAQVLSFLAAFGLVMAIPAILIGAIGASTDWNQTAYGLPDP-----KTTEEADMILP	321
		:: :   : : : :   : :   :   : :	
Db	308	SHVKAGCILCGYLKVMMPFLIVMMGMVSRILYTDKVACVVPSECERYCGTRVGCTNIAFP	367
Qy	322	IVLQYLCVPYISFFGLGAVSAVMSSADSSILSASSMFARNIYQLSFRQNASDKEIVWVM	381
		::     :   :   :             :   :   :	
Db	368	TLVVELMPNGLRGLMLSVMMASLMSLTSIFNSASTLFTMDIY-TKIRKKASEKELMIAG	426
Qy	382	RI-TVFVFGASATAMALLTKTVYG--LWYLSSDLVYI--VIFPQLLCVLFVKGNTYGA	436
		: :   :   : :   :   :   :   :	
Db	427	RLFMLFLIGISIAWVPIVQSAQSQQLFDYIQSITSYLGPPIAAVFLLAIFWKRVNEPGAF	486
Qy	437	AGYVSGFLRLI-----TG-----GEPYLYLQPLIFYPGYPPDDNGIY	473
		:                 :	
Db	487	WGLVLGFLIGISRMITEFAYGTGSCMEPSNCPTIICGVHYLYFAIILF-----	534
Qy	474	NQKFPFKTLAMVTSFLTNICISYLAHYLFESGTLPPKLDVFDVAVVA-RHSEENMDKTI	532
		:   :   :   :   : : :   :	
Db	535	-----VISIITVVVVSFLTQPI-----PDVHLYRLCWSLRNSKE-----	568
Qy	533	KNENIKLD--ELALVKPRQSMTLSTFTNKEAF-----LDVDSSPEGSGTED	577

Db 569 --ERIDLDAGEEDIQEAPEEATDTEVPKKKKGFFRRAYDLFCGLDQDKGPKMTKEEE 623

RESULT 2

SL52\_RAT

ID SL52\_RAT STANDARD; PRT; 670 AA.  
AC P53792;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Sodium/glucose cotransporter 2 (Na+)/glucose cotransporter 2)  
DE (Low affinity sodium-glucose cotransporter).  
GN SLC5A2 OR SGLT2.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Sprague-Dawley; TISSUE=Kidney;  
RX MEDLINE=96094332; PubMed=7493971;  
RA You G., Lee W.-S., Barros E.J.G., Kanai Y., Huo T.-L., Khawaja S.,  
RA Wells R.G., Nigam S.K., Hediger M.A.;  
RT "Molecular characteristics of Na(+)-coupled glucose transporters in  
RT adult and embryonic rat kidney."  
RL J. Biol. Chem. 270:29365-29371(1995).  
CC -!- FUNCTION: Sodium-dependent glucose transporter. Has a Na+ to  
CC glucose coupling ratio of 1:1.  
CC -!- FUNCTION: Efficient substrate transport in mammalian kidney is  
CC provided by the concerted action of a low affinity high capacity  
CC and a high affinity low capacity Na(+)/glucose cotransporter  
CC arranged in series along kidney proximal tubules.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -!- TISSUE SPECIFICITY: Kidney, in proximal tubule S1 segments.  
CC -!- DEVELOPMENTAL STAGE: Appears on embryonic day 17 and gradually  
CC increases until day 19. Decreases between day 19 and birth.  
CC -!- PTM: GLYCOSYLATED AT A SINGLE SITE.  
CC -!- SIMILARITY: BELONGS TO THE SODIUM:SOLUTE SYMPORTER FAMILY (SSF).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U29881; AAC52325.1; -.  
DR InterPro; IPR001734; Na/solut\_symport.  
DR Pfam; PF00474; SSF; 1.  
DR TIGRFAMs; TIGR00813; sss; 1.  
DR PROSITE; PS00456; NA\_SOLUT\_SYMP\_1; 1.  
DR PROSITE; PS00457; NA\_SOLUT\_SYMP\_2; 1.  
DR PROSITE; PS50283; NA\_SOLUT\_SYMP\_3; 1.  
KW Transport; Sugar transport; Transmembrane; Sodium transport; Symport;  
KW Glycoprotein.

FT	DOMAIN	1	23	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	24	42	POTENTIAL.
FT	DOMAIN	43	59	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	60	80	POTENTIAL.
FT	DOMAIN	81	100	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	101	121	POTENTIAL.
FT	DOMAIN	122	166	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	167	187	POTENTIAL.
FT	DOMAIN	188	203	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	204	224	POTENTIAL.
FT	DOMAIN	225	268	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	269	289	POTENTIAL.
FT	DOMAIN	290	312	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	313	332	POTENTIAL.
FT	DOMAIN	333	421	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	422	441	POTENTIAL.
FT	DOMAIN	442	453	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	454	474	POTENTIAL.
FT	DOMAIN	475	524	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	525	545	POTENTIAL.
FT	DOMAIN	546	648	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	649	669	POTENTIAL.
FT	CARBOHYD	248	248	N-LINKED (GLCNAC... ) (PROBABLE).
SQ	SEQUENCE	670 AA;	72961 MW;	0609562861618BB3 CRC64;

Query Match 10.4%; Score 308; DB 1; Length 670;  
 Best Local Similarity 23.3%; Pred. No. 2.9e-14;  
 Matches 148; Conservative 95; Mismatches 209; Indels 184; Gaps 28;

Qy	8	LIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGY	67
		::   ::  ::   ::        : :     :   ::   :     :	
Db	24	ILVIAAYFLLVIGVGLWSMFRT-NRGTV----GGYFLAGRSMVWVPGASLFASNIGSGH	78
Qy	68	INGTAEAVYVPGYGLAWAQAPIGYS-----LSLILGGLFFAKPMRSKGYVTMLDPFQQIY	122
		: :   :     : :   :	
Db	79	FVGLA-----GTGAASGLAVAGFEWNALFVVLGWLFPVYL-TAGVITM-----PQYL	127
Qy	123	GKRMGG-----LLFIPALMGEMFWAAAI--SALGATISVIIDVDMHISVIIS	168
		:   : :   :         :	
Db	128	RKRFGGRRIRLYLSVLSFLYIFTKISVDMFSGAVFIQQALGWN-----YASVIAL	179
Qy	169	ALIATLYTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKY----	224
		:    :    :            :   :   :	
Db	180	LGITMIYTVTGGLAALMYTDTVQTFVILAGAFILTGYAFHEVG----GYSGLFDKYLGAV	235
Qy	225	-----QKPWLGTVDSSSEVYSWLDSFLLL---MLGGIPW-----QAY	257
		:   :   :   :    :    :    :	
Db	236	TSLTVSKDPAVGNISSTCYQPRPDSYHLLRDPVTGGLPWPALLLGLTIVSGWHWCSDQVI	295
Qy	258	FQRLSSSSSATYQ----VLSFLAAFGCLVMAIPAILIGAIGASTDWNQTAYGLPDPKTT	313
		: :   : : : :   :   : : : :	
Db	296	VQRCLAGKNLTHIKAGCILCGYLKLMFPMFLMVMPGMI-----SRILY--PD----	339
Qy	314	EEADMILPIVLQYLC-----PVYISFFGLGAVSAAVMSSADSSILS	354
		: :     : :   :   :   :	
Db	340	-EVACVVPEVCKRVCGTEVGCNSIAYPRLVVKLMPNGLRGLMLAVMLAALMSSLASIFNS	398



QY 355 ASSMFARNIYQLSFRQNASDKEIVVWMRITVFVFGASATAMALLTKTVYG---LWYLSSD 411  
 :|::| :|| | | |:|:: | |: | : | : : | : |  
 Db 399 SSTLFTMDIY-TRLRPRAGDRELLLVGRLWVVFIVAVSVAWLPVVQAAQGGQLFDYIQSV 457  
 QY 412 LVYIV--IFPQLLCVLFVKGTNTYGAAGYVSGFLRLI-----TGG--EP----- 452  
 |: : : ||| | || | : || : : || |  
 Db 458 SSYLAPPVSAVFLALFVPRVNEKGAFWGLIGLLMGLARLIPEFFFGTGSCVRPSACPA 517  
 QY 453 -----YLYLQPLIFYPGYPPDDNGIYNQKFPFKTLAMVTSFLTNICISYLAKYLFESGT 506  
 ||| :|: | ||| : : | :| : :|  
 Db 518 IFCRVHYLYFAIILFFCS-----GFLTAL-ISRCTAPIPQKHLHRLVFS--- 560  
 QY 507 LPPKLDVFDVAVVARHSEENMDKTILVKENIKLDEL 542  
 |||:| :|:: :||  
 Db 561 -----LRHSKE-----EREDLDAEEL 576

# RESULT 3

## SL51\_HUMAN

ID SL51\_HUMAN STANDARD; PRT; 664 AA.  
 AC P13866;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Sodium/glucose cotransporter 1 (Na+)/glucose cotransporter 1)  
 DE (High affinity sodium-glucose cotransporter).  
 GN SLC5A1 OR SGLT1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89345544; PubMed=2490366;  
 RA Hediger M.A., Turk E., Wright E.M.;  
 RT "Homology of the human intestinal Na+/glucose and Escherichia coli  
 RT Na+/proline cotransporters.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 86:5748-5752(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Swan M.;  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20057165; PubMed=10591208;  
 RA Dunham I., Hunt A.R., Collins J.E., Bruskiewich R., Beare D.M.,  
 RA Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.K.,  
 RA Bagguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,  
 RA Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,  
 RA Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,  
 RA Clegg S.M., Copley V.E., Cole C.G., Collier R.E., Connor R.,  
 RA Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,  
 RA Dhami P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,  
 RA Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,  
 RA Gilbert J.G.R., Goward M.E., Grafham D.V., Griffiths M.N.D., Hall C.,  
 RA Hall R.E., Hall-Tamlyn G., Heathcote R.W., Ho S., Holmes S.,  
 RA Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,

RA Laird G.K., Langford C.F., Leversha M.A., Lloyd C., Lloyd D.M.,  
 RA Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., Mccann O.T.,  
 RA McClay J., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J.,  
 RA Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,  
 RA Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L., Ross M.T.,  
 RA Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,  
 RA Soderlund C., Spragon L., Steward C.A., Sulston J.E., Swann R.M.,  
 RA Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,  
 RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,  
 RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,  
 RA Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,  
 RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyama S.,  
 RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,  
 RA Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,  
 RA Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Pan H.,  
 RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaul S., Sloan D., Song L.,  
 RA Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,  
 RA Zhan M., Zhang G., Chisoe S., Murray J., Miller N., Minx P.,  
 RA Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,  
 RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,  
 RA Hinds K., Kemp K., Latreille P., Layman D., Ozersky P., Rohlfing T.,  
 RA Scheet P., Walker C., Wamsley A., Wohldmann P., Pepin K., Nelson J.,  
 RA Korf I., Bedell J.A., Hillier L.W., Mardis E., Waterston R.,  
 RA Wilson R., Emanuel B.S., Shaikh T., Kurahashi H., Saitta S.,  
 RA Budarf M.L., McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E.,  
 RA Edelmann L., Kim U.J., Shizuya H., Simon M.I., Dumanski J.P.,  
 RA Peyrard M., Kedra D., Seroussi E., Fransson I., Tapia I., Bruder C.E.,  
 RA O'Brien K.P., Wilkinson P., Bodenteich A., Hartman K., Hu X.,  
 RA Khan A.S., Lane L., Tilahun Y., Wright H.;

RT "The DNA sequence of human chromosome 22.";

RL Nature 402:489-495(1999).

RN [4]

RP VARIANT GGM ASN-28.

RX MEDLINE=91179516; PubMed=2008213;

RA Turk E., Zabel B., Mundlos S., Dyer J., Wright E.M.;

RT "Glucose/galactose malabsorption caused by a defect in the

RT Na<sup>+</sup>/glucose cotransporter.";

RL Nature 350:354-356(1991).

RN [5]

RP VARIANT GGM GLY-28.

RX MEDLINE=94253082; PubMed=8195156;

RA Turk E., Martin M.G., Wright E.M.;

RT "Structure of the human Na<sup>+</sup>/glucose cotransporter gene SGLT1.";

RL J. Biol. Chem. 269:15204-15209(1994).

CC -!- FUNCTION: Actively transports glucose into cells by Na<sup>+</sup>  
 CC co-transport with a Na<sup>+</sup> to glucose coupling ratio of 2:1.

CC -!- FUNCTION: Efficient substrate transport in mammalian kidney is  
 CC provided by the concerted action of a low affinity high capacity  
 CC and a high affinity low capacity Na<sup>+</sup>/glucose cotransporter  
 CC arranged in series along kidney proximal tubules.

CC -!- SUBCELLULAR LOCATION: Integral membrane protein.

CC -!- TISSUE SPECIFICITY: Expressed mainly in intestine and kidney.

CC -!- DISEASE: Defects in SLC5A1 are the cause of congenital glucose-  
 CC galactose malabsorption (GGM) [MIM:606824]. GGM is an intestinal  
 CC monosaccharide transporter deficiency. It is an autosomal  
 CC recessive disorder manifesting itself within the first weeks of  
 CC life. It is characterized by severe diarrhea and dehydration which

CC are usually fatal unless glucose and galactose are eliminated from  
 CC the diet.  
 CC -!- SIMILARITY: BELONGS TO THE SODIUM:SOLUTE SYMPORTER FAMILY (SSF).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; L29339; AAB59448.1; -.  
 DR EMBL; L29328; AAB59448.1; JOINED.  
 DR EMBL; L29330; AAB59448.1; JOINED.  
 DR EMBL; L29329; AAB59448.1; JOINED.  
 DR EMBL; L29331; AAB59448.1; JOINED.  
 DR EMBL; L29332; AAB59448.1; JOINED.  
 DR EMBL; L29333; AAB59448.1; JOINED.  
 DR EMBL; L29334; AAB59448.1; JOINED.  
 DR EMBL; L29335; AAB59448.1; JOINED.  
 DR EMBL; L29336; AAB59448.1; JOINED.  
 DR EMBL; L29337; AAB59448.1; JOINED.  
 DR EMBL; L29338; AAB59448.1; JOINED.  
 DR EMBL; M24847; AAA60320.1; -.  
 DR EMBL; AL022321; -; NOT\_ANNOTATED\_CDS.  
 DR EMBL; Z83849; -; NOT\_ANNOTATED\_CDS.  
 DR EMBL; Z74021; -; NOT\_ANNOTATED\_CDS.  
 DR EMBL; Z80998; -; NOT\_ANNOTATED\_CDS.  
 DR EMBL; Z83839; -; NOT\_ANNOTATED\_CDS.  
 DR PIR; A33545; A33545.  
 DR Genew; HGNC:11036; SLC5A1.  
 DR MIM; 182380; -.  
 DR MIM; 606824; -.  
 DR GO; GO:0005887; C:integral to plasma membrane; TAS.  
 DR InterPro; IPR001734; Na/solut\_symport.  
 DR Pfam; PF00474; SSF; 1.  
 DR TIGRFAMs; TIGR00813; sss; 1.  
 DR PROSITE; PS00456; NA\_SOLUT\_SYMP\_1; 1.  
 DR PROSITE; PS00457; NA\_SOLUT\_SYMP\_2; 1.  
 DR PROSITE; PS0283; NA\_SOLUT\_SYMP\_3; 1.  
 KW Transport; Sugar transport; Transmembrane; Sodium transport; Symport;  
 KW Glycoprotein; Disease mutation.  
 FT DOMAIN 1 28 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 29 47 POTENTIAL.  
 FT DOMAIN 48 64 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 65 85 POTENTIAL.  
 FT DOMAIN 86 105 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 106 126 POTENTIAL.  
 FT DOMAIN 127 171 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 172 192 POTENTIAL.  
 FT DOMAIN 193 208 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 209 229 POTENTIAL.  
 FT DOMAIN 230 270 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 271 291 POTENTIAL.  
 FT DOMAIN 292 314 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 315 334 POTENTIAL.



```

Db      529 FAIILF-----AISFITIVVISLLTKPI-----PDVHLYR 558
QY      516 AV--VARHSEENMDKTILVKENIKLDELALVKPRQSMTLSTFTNKE 561
      :   || :| : : |||:   |::: : :   |:
Db      559 LCWSLRNSKEERID--LDAEEENIQ-----EGPKETIEIETQVPEKK 598

```

#### RESULT 4

##### SL51\_RAT

```

ID      SL51_RAT          STANDARD;          PRT;          665 AA.
AC      P53790; P97787;
DT      01-OCT-1996 (Rel. 34, Created)
DT      01-OCT-1996 (Rel. 34, Last sequence update)
DT      10-OCT-2003 (Rel. 42, Last annotation update)
DE      Sodium/glucose cotransporter 1 (Na+)/glucose cotransporter 1)
DE      (High affinity sodium-glucose cotransporter).
GN      SLC5A1 OR SGLT1.
OS      Rattus norvegicus (Rat).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX      NCBI_TaxID=10116;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=Sprague-Dawley; TISSUE=Kidney;
RX      MEDLINE=94216314; PubMed=8163506;
RA      Lee W.S., Kanai Y., Wells R.G., Hediger M.A.;
RT      "The high affinity Na+/glucose cotransporter. Re-evaluation of
RT      function and distribution of expression.";
RL      J. Biol. Chem. 269:12032-12039(1994).
RN      [2]
RP      SEQUENCE FROM N.A.
RA      Kasahara M., Mori K.;
RL      Submitted (MAY-1993) to the EMBL/GenBank/DDBJ databases.
RN      [3]
RP      SEQUENCE FROM N.A.
RC      STRAIN=Sprague-Dawley; TISSUE=Jejunum;
RA      Aoshima H., Yokoyama T., Tanizaki J., Izu H., Yamada M.;
RT      "The sugar specificity of Na/glucose cotransporter from rat jejunum.";
RL      Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
CC      -!- FUNCTION: Actively transports glucose into cells by Na(+) co-
CC      transport with a Na(+) to glucose coupling ratio of 2:1.
CC      -!- FUNCTION: Efficient substrate transport in mammalian kidney is
CC      provided by the concerted action of a low affinity high capacity
CC      and a high affinity low capacity Na(+)/glucose cotransporter
CC      arranged in series along kidney proximal tubules.
CC      -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC      -!- DEVELOPMENTAL STAGE: Appears on embryonic day 18 and gradually
CC      increases until birth.
CC      -!- SIMILARITY: BELONGS TO THE SODIUM:SOLUTE SYMPORTER FAMILY (SSF).
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).

```



```

Db      189 ALYTITGGLAAVIYDTLQTAIMLVGSFILTGFAPREV-----GYEAFMDKYMKAIPTLV 244
Qy      230 --GTVD--SSEVYS-WLDSFLLL---MLGGIPW-----QAYFQRVLSS 264
      | : | | : || : : | : || | || :
Db      245 SDGNITVKEECYTPRADSFHIFRDPITGMPWPGLIFGLSILALWYWCTDQVIVQRCLSA 304
Qy      265 SSATYAQVLSFLAAFGCLVMAIPAILIGAIGASTDWNQTAYGLPDP-----KTTEEADM 318
      : :: : | : | : : : | : | : :
Db      305 KNMSHVKAGCTLGCGYLKLLPMFLMVMPGMISRILYTDKIACVLPSECKKYCGTPVGCTNI 364
Qy      319 ILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFARNIYQLSFRQNASDKEIV 378
      | :: | | : | : | : || | | || : | : | : | :
Db      365 AYPTLVVELMPNGLRGLMLSVMMASLMSLTSLFNSASTLFTMDIY-TKIRKGASEKELM 423
Qy      379 WVMRITVFV-FGASATAMALLTKTVYG--LWYLSSDLVYI--VIFPQLLCVLFVKGTNTY 433
      | : : | | | : : : | : | : | : | : | : |
Db      424 IAGRLFILVLIGISIAWVPIVQSAQSGQLFDYIQSITSYLGPPIAAVFLLAIFCKRVNEP 483
Qy      434 GAVAGYVSGLEFLRI-----TG-----GEPYLYLQPLIFYPGYPPDDN 470
      || | : | : | | | | | | | | : |
Db      484 GAFWGLILGFLIGISRMITFAYGTGSCMEPSNCPKIICGVHYLYFAILF----- 534
Qy      471 GIYNQKFPFKTLAMVTSFLTNICISYLAHYLFESGTLPPKLDVFDVAV--VARHSEENMDK 528
      | : | : || | | : | : : : : : | : |
Db      535 -----AISVVTVLVISLLTKPI-----PDVHLYRLCWSLRNSTEERID- 572
Qy      529 TILVKNENIKLDELALVKPRQSMTLSSTFTNKE-----AFLDVDSSPEGSGTED 577
      | | : | | : : : : | | | | : : | :
Db      573 --LDAGEEEFVEE-----DPKDTIEIDAEAPQKEKGCFRKAYDLFCGLDQDKGPKMTKEEE 626

```

# RESULT 5

## SL54\_PIG

```

ID SL54_PIG STANDARD; PRT; 660 AA.
AC P31636;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Low affinity sodium-glucose cotransporter (Sodium/glucose
DE cotransporter 3) (Na+)/glucose cotransporter 3).
GN SLC5A4 OR SGLT3 OR SAAT1.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=93131881; PubMed=8420925;
RA Kong C.-T., Yet S.-F., Lever J.E.;
RT "Cloning and expression of a mammalian Na+/amino acid cotransporter
RT with sequence similarity to Na+/glucose cotransporters.";
RL J. Biol. Chem. 268:1509-1512(1993).
RN [2]
RP FUNCTION.
RX MEDLINE=94357885; PubMed=8077195;
RA McKenzie B., Panayotova-Heiermann M., Loo D.D.F., Lever J.E.,

```

RA Wright E.M.;  
 RT "SAAT1 is a low affinity Na<sup>+</sup>/glucose cotransporter and not an amino  
 RT acid transporter. A reinterpretation."  
 RL J. Biol. Chem. 269:22488-22491(1994).  
 CC -!- FUNCTION: Sodium-dependent glucose transporter. Has a Na<sup>+</sup> to  
 CC glucose coupling ratio of 1:1.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -!- TISSUE SPECIFICITY: KIDNEY, INTESTINE, LIVER, SKELETAL MUSCLE,  
 CC AND SPLEEN.  
 CC -!- SIMILARITY: BELONGS TO THE SODIUM:SOLUTE SYMPORTER FAMILY (SSF).  
 CC -!- CAUTION: Was originally (Ref.1) thought to be a sodium/neutral  
 CC amino acid cotransporter (system a neutral amino acid transporter)  
 CC responsible for the sodium-dependent intake of neutral amino acids  
 CC such as alanine, glycine, serine, cysteine, and proline.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; L02900; AAC37325.1; -.  
 DR PIR; A44432; A44432.  
 DR InterPro; IPR001734; Na/solut\_sympot.  
 DR Pfam; PF00474; SSF; 1.  
 DR TIGRFAMs; TIGR00813; sss; 1.  
 DR PROSITE; PS00456; NA\_SOLUT\_SYMP\_1; 1.  
 DR PROSITE; PS00457; NA\_SOLUT\_SYMP\_2; 1.  
 DR PROSITE; PS50283; NA\_SOLUT\_SYMP\_3; 1.  
 KW Transport; Sugar transport; Transmembrane; Sodium transport; Symport;  
 KW Glycoprotein.  
 FT DOMAIN 1 28 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 29 47 POTENTIAL.  
 FT DOMAIN 48 64 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 65 85 POTENTIAL.  
 FT DOMAIN 86 105 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 106 126 POTENTIAL.  
 FT DOMAIN 127 171 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 172 191 POTENTIAL.  
 FT DOMAIN 192 208 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 209 229 POTENTIAL.  
 FT DOMAIN 230 270 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 271 291 POTENTIAL.  
 FT DOMAIN 292 314 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 315 334 POTENTIAL.  
 FT DOMAIN 335 423 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 424 443 POTENTIAL.  
 FT DOMAIN 444 455 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 456 476 POTENTIAL.  
 FT DOMAIN 477 526 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 527 547 POTENTIAL.  
 FT DOMAIN 548 638 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 639 659 POTENTIAL.  
 FT CARBOHYD 248 248 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 660 AA; 72745 MW; 38616367F8F18F1A CRC64;



Query Match 10.2%; Score 303.5; DB 1; Length 660;  
 Best Local Similarity 23.2%; Pred. No. 5.9e-14;  
 Matches 141; Conservative 103; Mismatches 230; Indels 135; Gaps 26;

```

Qy      11 IIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGYING 70
        |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      32 IVIYFVVVMAVGLWAMLRT-NRGTV-----GGFFLAGRDVTWWPMGASLFASNIGSGHFVG 86

Qy      71 TAEAVYVPGYGLA---WAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQIY-GKRM 126
        |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      87 LAGTGAASGIAIAAFEW-----NALLLLVLGWFFVPIYIKAGVMTMPEYLRKRFGGKRL 141

Qy     127 GGLLFIPAL-----MGEMFWAAAFSALGATISVIIDVDMHISVIISALIATLYTLVG 179
        |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db     142 QIYLSILSLFICVALRISSDIFSGAIF-----IKLALGLDLYLAIFSLLAITAIYTTITG 195

Qy     180 GLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQK--PWLGTVD---- 233
        |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db     196 GLASVIYTDTLQTIIMLIGSFILMGFAF----VEVGGYESFTEKYMNAIPTIVEGDNLT 251

Qy     234 SSEVYS-WLDSFLLL---MLGGIPW-----QAYFQRLSSSSATYAQ 271
        |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db     252 SPKCYTPQGDSFHIFRDAVTGDI PWPGMIFGMTVVAAWYWCTDQVIVQRCLSGKDMSHVK 311

Qy     272 VLSFLAAFGCLVMAIPAILIGAIGASTDWNQTAYGLPDPKT----TEE--ADMILPIVLQ 325
        : : |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db     312 AACIMCGYLKLLPMFLMVMPGMISRILYTEKVACVVPSECVKHCGETVGCSNYAYPLLVM 371

Qy     326 YLCPVYISFFGLGAVSAAMSSADSSILSASSMFARNIYQLSFRQNASDKEIVWVMRITV 385
        || : |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db     372 ELMPSGLRGLMLSVMLASLMSSLTSIFNSASTLFTMDLY-TKIRKQASEKELLIAGRLFI 430

Qy     386 FVFGASATAMALLTKTVYG---LWYLSSDLVYI--VIFPQLLCVLFVKGTNTYGA----V 436
        : : |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db     431 ILLIVISIVWVPLVQVAQNGQLFHYIESISSYLGPPIAAVFLLAIFCKRVNEQGAFWGLI 490

Qy     437 AGYVSGL-----FLRITG-----GEPYLYLQPLIFYPGYYPDDNGIYNQKF 477
        |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db     491 IGFV MGLIRMIAEFVYGTGSCLAASNCPQIICGVHYLYFALILFF----- 535

Qy     478 PFKTLAMVTSFLTNICISYLAK-----YLFE-----SGTLPPKLDVFDVAVARH----- 521
        | : |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db     536 -----VSILVVLAI SLTKPIPDVHLYRLCWALRNSTEERIDL-DAEKRHEEAHDG 586

Qy     522 -SEENMDKT 529
        |::|::|
Db     587 VDEDNPEET 595
  
```

# RESULT 6

SL52\_RABIT

ID SL52\_RABIT STANDARD; PRT; 672 AA.

AC P26430;

DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Sodium/nucleoside cotransporter (Na(+)/nucleoside cotransporter).  
 GN SLC5A2 OR SNST1.  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_TaxID=9986;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Kidney;  
 RX MEDLINE=92156077; PubMed=1740408;  
 RA Pajor A.M., Wright E.M.;  
 RT "Cloning and functional expression of a mammalian Na<sup>+</sup>/nucleoside  
 RT cotransporter. A member of the SGLT family."  
 RL J. Biol. Chem. 267:3557-3560(1992).  
 CC -!- FUNCTION: Actively transports uridine into cells by Na<sup>+</sup>  
 CC co-transport. May play a role in reabsorption of nucleosides from  
 CC glomerular filtrate by the proximal tubule in kidney, and in the  
 CC regulation of cardiac contractility by adenosine.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -!- TISSUE SPECIFICITY: More abundant in heart than in kidney, where  
 CC it is absent from the outer cortex.  
 CC -!- SIMILARITY: BELONGS TO THE SODIUM:SOLUTE SYMPORTER FAMILY (SSF).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M84020; AAA31421.1; -.  
 DR PIR; A42251; A42251.  
 DR InterPro; IPR001734; Na/solut\_symport.  
 DR Pfam; PF00474; SSF; 1.  
 DR TIGRFAMs; TIGR00813; sss; 1.  
 DR PROSITE; PS00456; NA\_SOLUT\_SYMP\_1; 1.  
 DR PROSITE; PS00457; NA\_SOLUT\_SYMP\_2; 1.  
 DR PROSITE; PS50283; NA\_SOLUT\_SYMP\_3; 1.  
 KW Transport; Transmembrane; Sodium transport; Symport; Glycoprotein.  
 FT DOMAIN 1 25 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 26 44 POTENTIAL.  
 FT DOMAIN 45 61 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 62 82 POTENTIAL.  
 FT DOMAIN 83 102 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 103 123 POTENTIAL.  
 FT DOMAIN 124 168 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 169 188 POTENTIAL.  
 FT DOMAIN 189 205 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 206 226 POTENTIAL.  
 FT DOMAIN 227 270 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 271 291 POTENTIAL.  
 FT DOMAIN 292 314 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 315 334 POTENTIAL.  
 FT DOMAIN 335 422 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 423 442 POTENTIAL.  
 FT DOMAIN 443 454 CYTOPLASMIC (POTENTIAL).

FT	TRANSMEM	455	475	POTENTIAL.
FT	DOMAIN	476	525	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	526	546	POTENTIAL.
FT	DOMAIN	547	650	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	651	671	POTENTIAL.
FT	CARBOHYD	250	250	N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ	SEQUENCE	672 AA;	73161 MW;	E2D987B03B9C57B4 CRC64;

Query Match 10.0%; Score 298; DB 1; Length 672;  
 Best Local Similarity 25.0%; Pred. No. 1.5e-13;  
 Matches 153; Conservative 89; Mismatches 232; Indels 138; Gaps 25;

Qy	9	IAII-VFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGY	67
		:  : :  : :  :  :  :  :  :  :	
Db	26	IAVIAAYFLLVIGVGLWSMCRN-NRGTV----GGYFLAGRSMVWVPGASLFASNIGSGH	80
Qy	68	INGTAEAVYVPGYGLAWAQAPIGYSLS----LILGGLFFAKPMRSKGYVTMLDPFQQIYG	123
		: : : :   :   :   :	
Db	81	FVGLA-----GTGAANGLAVAGFEWNALFVLLLGWLFAPVYLTAGVITM----PQYLR	130
Qy	124	KRMGG-----LLFIPALMGEMFWAAAI--SALGATISVIIDVDMHISVIISA	169
		:  : : :       :	
Db	131	KRFGGHRIRLYLSVLSFLYIFTKISVDMFSGAVFIQQALGWNI-----YASVIALL	182
Qy	170	LIATLYTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKY-----	224
		: :  :  :  :  :  :  :  :  :  :	
Db	183	GITMVYTVTGGLAALMYTDTVQTFVIIAGAFILTGYAFHEVG----GYSGLFDKYMGAMT	238
Qy	225	-----QKPWLGTVDSSSEVYSWLDSELLL---MLGGIPW-----QAYF	258
		:   :  :  :  :  :  :  :  :	
Db	239	SLTVSEDPVAGNIISSCYRPRPDSYHLLRDPVTGDLFPWALLLGLTIVSGWYWCSDQVIV	298
Qy	259	QRVLSSSSATYAQVLSFLAAFGLVMAIPAILIGAIGASTDWNQTAYGLPDPKT----TE	314
		:  :  :  :  :  :  :  :  :  :	
Db	299	QRCLAGRNLTHIKAGCILCGYLKLTMPFLMVMPGMISRIYPDEVACVAPEVCKRVCSTE	358
Qy	315	E--ADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFARNIYQLSFRQNA	372
		: :  :  :  :  :  :  :  :  :  :  :	
Db	359	VGCSNIAYPRLVVKLMPNGLRGLMLAVMLAALMSSLASIFNSSSTLFTMDIYTL--RPRA	416
Qy	373	SDKEIVWVMRITVFVFGASATAMALLTKTVYG---LWYLSSDLVYIV--IFPQLLCVLFV	427
		:  :  :  :  :  :  :  :  :  :  :	
Db	417	GEGELLVGRWLWVVFIVAVSVAWLPVVQAAQGGQLFDYIQSVSSYLAPPVSAVFVVALFV	476
Qy	428	KGTNTYGAVAGYVSGFLRITGGEPYLYLQPLIFYPGYPPDDNGIYNQKFPFKTLAMV--	485
		:  :  :  :  :  :  :  :  :	
Db	477	PRVNEKGAFWGLIGLLMGLARLIP-----EFSFGTGSCVRP	513
Qy	486	---TSFLTNICISYLAKYLFE-SG-----TLP-PKLDVFDVAVVA-RHSEENMDKTI	530
		:  :  :  :  :  :  :  :  :  :  :	
Db	514	SACPAFLCRVHYLYFAIVLFFCSGLLIIIVSLCTAPIPRKHLHRLVFSLRHSKE-----	567
Qy	531	LVKNENIKLDEL	542
		:  :  :	
Db	568	--EREDLDADEL	577

RESULT 7

SL54\_HUMAN

ID SL54\_HUMAN STANDARD; PRT; 659 AA.  
AC Q9NY91; O15279;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Low affinity sodium-glucose cotransporter (Sodium/glucose  
DE cotransporter 3) (Na(+)/glucose cotransporter 3).  
GN SLC5A4 OR SAAT1 OR SGLT2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Small intestine;  
RA Gorboulev V., Baumgarten K., Veyhl M., Koepsell H.;  
RT "The molecular cloning and functional characterization of the human  
RT SGLT2 transporter.";  
RL Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20057165; PubMed=10591208;  
RA Dunham I., Hunt A.R., Collins J.E., Bruskiewich R., Beare D.M.,  
RA Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.K.,  
RA Bagguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,  
RA Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,  
RA Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,  
RA Clegg S.M., Cogley V.E., Cole C.G., Collier R.E., Connor R.,  
RA Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,  
RA Dhami P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,  
RA Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,  
RA Gilbert J.G.R., Goward M.E., Grafham D.V., Griffiths M.N.D., Hall C.,  
RA Hall R.E., Hall-Tamlyn G., Heathcott R.W., Ho S., Holmes S.,  
RA Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,  
RA Laird G.K., Langford C.F., Leversha M.A., Lloyd C., Lloyd D.M.,  
RA Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., Mccann O.T.,  
RA Mcclay J., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J.,  
RA Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,  
RA Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L., Ross M.T.,  
RA Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,  
RA Soderlund C., Spragon L., Steward C.A., Sulston J.E., Swann R.M.,  
RA Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,  
RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,  
RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,  
RA Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,  
RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyama S.,  
RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,  
RA Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,  
RA Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Pan H.,  
RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,  
RA Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,  
RA Zhan M., Zhang G., Chissoe S., Murray J., Miller N., Minx P.,  
RA Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,  
RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,  
RA Hinds K., Kemp K., Latreille P., Layman D., Ozersky P., Rohlfling T.,

RA Scheet P., Walker C., Wamsley A., Wohldmann P., Pepin K., Nelson J.,  
 RA Korf I., Bedell J.A., Hillier L.W., Mardis E., Waterston R.,  
 RA Wilson R., Emanuel B.S., Shaikh T., Kurahashi H., Saitta S.,  
 RA Budarf M.L., McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E.,  
 RA Edelmann L., Kim U.J., Shizuya H., Simon M.I., Dumanski J.P.,  
 RA Peyrard M., Kedra D., Seroussi E., Fransson I., Tapia I., Bruder C.E.,  
 RA O'Brien K.P., Wilkinson P., Bodenteich A., Hartman K., Hu X.,  
 RA Khan A.S., Lane L., Tilahun Y., Wright H.;  
 RT "The DNA sequence of human chromosome 22."  
 RL Nature 402:489-495(1999).  
 RN [3]  
 RP SEQUENCE OF 73-247 FROM N.A.  
 RC TISSUE=Brain;  
 RA Poppe R., Koepsell H.;  
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: Sodium-dependent glucose transporter (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -!- SIMILARITY: BELONGS TO THE SODIUM:SOLUTE SYMPORTER FAMILY (SSF).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AJ133127; CAB81772.1; -.  
 DR EMBL; AL008723; CAB51758.1; -.  
 DR EMBL; U41897; AAB61732.1; -.  
 DR Genew; HGNC:11039; SLC5A4.  
 DR InterPro; IPR001734; Na/solut\_symport.  
 DR Pfam; PF00474; SSF; 1.  
 DR TIGRFAMs; TIGR00813; sss; 1.  
 DR PROSITE; PS00456; NA\_SOLUT\_SYMP\_1; 1.  
 DR PROSITE; PS00457; NA\_SOLUT\_SYMP\_2; 1.  
 DR PROSITE; PS50283; NA\_SOLUT\_SYMP\_3; 1.  
 KW Transport; Sugar transport; Transmembrane; Sodium transport; Symport;  
 KW Glycoprotein.  
 FT DOMAIN 1 28 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 29 47 POTENTIAL.  
 FT DOMAIN 48 64 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 65 85 POTENTIAL.  
 FT DOMAIN 86 105 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 106 126 POTENTIAL.  
 FT DOMAIN 127 171 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 172 191 POTENTIAL.  
 FT DOMAIN 192 208 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 209 229 POTENTIAL.  
 FT DOMAIN 230 270 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 271 291 POTENTIAL.  
 FT DOMAIN 292 314 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 315 334 POTENTIAL.  
 FT DOMAIN 335 423 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 424 443 POTENTIAL.  
 FT DOMAIN 444 455 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 456 476 POTENTIAL.

FT	DOMAIN	477	526	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	527	547	POTENTIAL.
FT	DOMAIN	548	637	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	638	658	POTENTIAL.
FT	CARBOHYD	248	248	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CONFLICT	76	76	A -> V (IN REF. 3).
FT	CONFLICT	106	106	S -> P (IN REF. 3).
FT	CONFLICT	243	243	V -> I (IN REF. 3).
SQ	SEQUENCE	659 AA; 72455 MW; F8A34AED648B523A CRC64;		

Query Match 9.9%; Score 294; DB 1; Length 659;  
 Best Local Similarity 22.5%; Pred. No. 2.7e-13;  
 Matches 135; Conservative 97; Mismatches 225; Indels 144; Gaps 22;

Qy	11	IIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGYING	70
		::: :::   :        :       :   :   :   :	
Db	32	IVIYFLVMAVGLWAMLKT-NRGTI----GGFFLAGRDMAWWPMGASLFASNIGSNHYVG	86
Qy	71	TAEAVYVPGYGLAWAQAPIGYS-----LSLILGGLFFAKPMRSKGYVTMLDPFQQIYGKR	125
		: : :       :   :     :     :	
Db	87	LA-----GTGAASGVATVTFEWTSSVMLLILGWIFVPIYIKS-GVMTM----PEYLKKR	135
Qy	126	MGG-----LLFIPALMGEMFWAAAI FSALGATISVIIDVDMHISVIISALIAT	173
		: :   : :         : : :   : : :   :	
Db	136	FGGERLQVYLSILSLFICVLLISADIFAGAIF-----IKLALGLDLYLAIFILLAMTA	189
Qy	174	LYTLVGGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVD	233
		:               :   : :   :   :   :   :   :	
Db	190	VYTTTGGLASVIYTDTLQTIIMLIGSFILMGFAFNEVG----GYESFTEKYVNATPSVVE	245
Qy	234	-----SSEVYS-WLDSFLLL--MLGGIPW-----QAYFQRVLSSS	265
		:   :     : :	
Db	246	GDNLTIASCYTPRADSFHIFRDAVTGDIWPGGIIFGMPITALWYWCTNQVIVQRCLCGK	305
Qy	266	SATYAQVLSFLAAFGLVMAIPAILIGAIGASTDWNQTAYGLP-----DPKTTEEA	316
		: : : :   :   : :     :   :	
Db	306	DMSHVKAACIMCAYLKLLPMFLMVMPGMISRILYTDMVACVVPSECVKHCGVDVGCTNYA	365
Qy	317	DMILPIVLQYLCVPYISFFGLGAVSAAVMSSADSSILSASSMFARNIYQLSFRQNASDKE	376
		: :     :   :   :             :   :   :	
Db	366	---YPTMVLELMPQGLRGLMLSVMLASLMSSLTSIFNSASTLFTIDLY-TKMRKQASEKE	421
Qy	377	IVWVMRITVFVFGASATAMALLTKTVYG---LWYLSSDLVYI--VIFPQLLCVLFVKGTN	431
		: :       : :   :   :     :   :   :	
Db	422	LLIAGRIFVLLTVVSIVWVPLVQVSQNGQLIHYTESISSYLGPPIAAVFVLAIFCKRVN	481
Qy	432	TYGAVAGYVSGLFL-----RITGGEPYLYLQPLIFYPGYYPD	468
		:     : :           : :	
Db	482	EQGAFWGLMVGLAMGLIRMITEFAYGTGSCCLAPSNCPKIICGVHYLYFSIVLFF-----	535
Qy	469	DNGIYNQKFPFKTLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDA--VVARHSEENM	526
		:       :   : :   : :   :	
Db	536	-----GSMLVTLGISLLTKPI-----PDVHLYRLCWVLRNSTEERI	571
Qy	527	D 527	
Db	572	D 572	

RESULT 8

SL51\_SHEEP

ID SL51\_SHEEP STANDARD; . PRT; 664 AA.  
AC P53791;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Sodium/glucose cotransporter 1 (Na+)/glucose cotransporter 1)  
DE (High affinity sodium-glucose cotransporter).  
GN SLC5A1 OR SGLT1.  
OS Ovis aries (Sheep).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;  
OC Bovidae; Caprinae; Ovis.  
OX NCBI\_TaxID=9940;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Jejunal mucosa;  
RX MEDLINE=96077158; PubMed=7492327;  
RA Tarpey P., Wood I.S., Shirazi-Beechey S.P., Beechey R.B.;  
RT "Amino acid sequence and the cellular location of the Na(+)-dependent  
RT D-glucose symporters (SGLT1) in the ovine enterocyte and the parotid  
RT acinar cell.";  
RL Biochem. J. 312:293-300(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Mammary gland;  
RX MEDLINE=98050042; PubMed=9388688;  
RA Shillingford J.M., Wood I.S., Shennan D.B., Shirazi-Beechey S.P.,  
RA Beechey R.B.;  
RT "Determination of the sequence of a mRNA from lactating sheep mammary  
RT gland that encodes a protein identical to the Na(+)-dependent glucose  
RT transporter (SGLT1).";  
RL Biochem. Soc. Trans. 25:467-467(1997).  
CC -!- FUNCTION: Actively transports glucose into cells by Na(+) co-  
CC transport with a Na(+) to glucose coupling ratio of 2:1.  
CC -!- FUNCTION: Efficient substrate transport in mammalian kidney is  
CC provided by the concerted action of a low affinity high capacity  
CC and a high affinity low capacity Na(+)/glucose cotransporter  
CC arranged in series along kidney proximal tubules.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -!- SIMILARITY: BELONGS TO THE SODIUM:SOLUTE SYMPORTER FAMILY (SSF).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X82411; CAA57809.1; -.  
DR EMBL; X82410; CAA57808.1; -.  
DR EMBL; AJ001026; CAA04483.1; -.  
DR PIR; S59637; S59637.





```

Qy      266 SATYAQVLSFLAAFGLVMAIPAILIGAIGASTDWNQTAYGLPDPKTTEE-----AD 317
      : :: :      : : | :      :: | |      : | : | | :      :
Db      306 NMSHVKAGCIMCGYMKLLPMFLMVMPGMISRILFTEKVACTV--PSECEKYCGTKVGCTN 363

Qy      318 MILPIVLQYLCVPYISFFGLGAVSAVMSSADSSILSASSMFARNIYQLSFRQNASDKEI 377
      : | :: | | :      | : |::||| |      |||::| :||      |: ||::||:
Db      364 IAYPTLVVELMPNGLRGLMLSVMLASIMSSLTSIFNSASTLFTMDIY-TKIRKKASEKEL 422

Qy      378 VWVMRITVFV-FGASATAMALLTKTVYG--LWYLSSDLVYI--VIFPQLLCVLFVKGTNT 432
      : | : : | | | : ::      |      | : |      | :      | : | | |
Db      423 MIAGRLFMLVLIGVSIAWVPIVQSAQSGQLFDYIQSITSYLGPPIAAVFLLAIFCKRVNE 482

Qy      433 YGAVAGYVSGFLFRI-----TG-----GEPYLYLQPLIF 461
      || | : | : :      ||      | ||| : :|
Db      483 PGAFWGLIIGFLIGVSRMITEFAYGTGSCMEPSNCPTIICGVHYLYFAIILF 534

```

# RESULT 9

## SGLT\_VIBPA

```

ID      SGLT_VIBPA      STANDARD;      PRT;      543 AA.
AC      P96169;
DT      30-MAY-2000 (Rel. 39, Created)
DT      30-MAY-2000 (Rel. 39, Last sequence update)
DT      28-FEB-2003 (Rel. 41, Last annotation update)
DE      Sodium/glucose cotransporter (Na+)/glucose symporter).
GN      SGLT.
OS      Vibrio parahaemolyticus.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC      Vibrionaceae; Vibrio.
OX      NCBI_TaxID=670;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=AQ3334;
RX      MEDLINE=96248401; PubMed=8652595;
RA      Sarker R.I., Okabe Y., Tsuda M., Tsuchiya T.;
RT      "Sequence of a Na+/glucose symporter gene and its flanking regions of
RT      Vibrio parahaemolyticus.";
RL      Biochim. Biophys. Acta 1281:1-4(1996).
RN      [2]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=20400508; PubMed=10835424;
RA      Turk E., Kim O., Le Coutre J., Whitelegge J.P., Eskandari S.,
RA      Lam J.T., Kreman M., Zampighi G., Faull K.F., Wright E.M.;
RT      "Molecular characterization of Vibrio parahaemolyticus vSGLT: a model
RT      for sodium-coupled sugar cotransporters.";
RL      J. Biol. Chem. 275:25711-25716(2000).
RN      [3]
RP      MASS SPECTROMETRY OF FORMYLATED FORM, AND REVISIONS TO N-TERMINUS.
RX      MEDLINE=20222957; PubMed=10757971;
RA      le Coutre J., Whitelegge J.P., Gross A., Turk E., Wright E.M.,
RA      Kaback H.R., Faull K.F.;
RT      "Proteomics on full-length membrane proteins using mass
RT      spectrometry.";
RL      Biochemistry 39:4237-4242(2000).
CC      -!- FUNCTION: ACTIVELY TRANSPORTS GLUCOSE INTO CELLS BY NA(+) CO-
CC      TRANSPORT (BY SIMILARITY).
CC      -!- SUBCELLULAR LOCATION: Integral membrane protein.

```

```

CC  -!- MASS SPECTROMETRY: MW=60680; METHOD=Electrospray.
CC  -!- SIMILARITY: BELONGS TO THE SODIUM:SOLUTE SYMPORTER FAMILY (SSF).
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; D78137; BAA11215.1; ALT_FRAME.
DR  EMBL; AF255301; AAF80602.1; -.
DR  InterPro; IPR001734; Na/solut_sympot.
DR  Pfam; PF00474; SSF; 1.
DR  TIGRFAMS; TIGR00813; sss; 1.
DR  PROSITE; PS00456; NA_SOLUT_SYMP_1; 1.
DR  PROSITE; PS00457; NA_SOLUT_SYMP_2; FALSE_NEG.
DR  PROSITE; PS50283; NA_SOLUT_SYMP_3; 1.
KW  Transport; Sugar transport; Transmembrane; Sodium transport; Sympot.
FT  TRANSMEM      10      30      POTENTIAL.
FT  TRANSMEM      45      65      POTENTIAL.
FT  TRANSMEM      79      99      POTENTIAL.
FT  TRANSMEM     129     149      POTENTIAL.
FT  TRANSMEM     156     176      POTENTIAL.
FT  TRANSMEM     193     213      POTENTIAL.
FT  TRANSMEM     246     266      POTENTIAL.
FT  TRANSMEM     287     307      POTENTIAL.
FT  TRANSMEM     345     365      POTENTIAL.
FT  TRANSMEM     401     421      POTENTIAL.
FT  TRANSMEM     427     447      POTENTIAL.
FT  TRANSMEM     455     475      POTENTIAL.
FT  TRANSMEM     483     503      POTENTIAL.
FT  TRANSMEM     523     543      POTENTIAL.
SQ  SEQUENCE      543 AA;  58874 MW;  61BE3F7E380BC32C CRC64;

```

```

Query Match          9.9%;  Score 293.5;  DB 1;  Length 543;
Best Local Similarity 25.1%;  Pred. No. 2.4e-13;
Matches 139;  Conservative 96;  Mismatches 198;  Indels 121;  Gaps 27;

```

```

Qy      4 HVEGLIAIVF--YLLILL-VGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGFTMTA 60
      |  |  |||  |:  ||:  ||:|  ::  ::  :|  :  |:  :  |  ::  |
Db      6 HGLSFIDIMVFAIYVAIIIGVGLWV---SRDKKGTQKSTEDYFLAGKSLPWWAVGASLIA 62

Qy     61 TWV-----GGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGY 111
      :  |  ||  ||  ||  ||  |  |  :||:  |:  |  ||
Db     63 ANISAEQFIGMSGSGYSIGLAIASY-----EWMSA-----ITLIIVGKYFLPIFIEKGI 111

Qy    112 VTMLDPFQQIYGKRMGGLLFIPALMGEMFWAAA-IFSAL-----GATISVIIDVDMHI 163
      |:  :  ::  :  |:  :  |  :  ||  :  ||  |  |  :  |:  :  :
Db    112 YTIPEFVEKRFNKKLKTILAV-----FWISLYIFVNLTSVLYLGGLALETILGIPLMY 164

Qy    164 SVIISALIATLYTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAK 223
      |:  :  ||  |  |:  :  ||  |  |:  ||:  |:  |  :  |  |  :  ||:  ||  |
Db    165 SILGLALFALVYSIYGGLSAVVWTDVIQVFFLVLG-----GFMTTYMAVSFIGGT----- 214

Qy    224 YQKPWLGTV-----DSSEVYSWLDSFLLMLGGIPW-----QAY 257

```

```

      | | | | | : | | : : : | | | |
Db      215 --DGWFAGVSKMVDAAPGHFEMILDQSNPQYMNLPG-IAVLIGGL-WVANLYYWGFNQYI 270

QY      258 FQRVLSSSSATYAQVLSFLAAGCLVMAIPAILIG-AIGASTDWNQTAYGLPDPKTTE-- 314
      || |:: | : || || |:: : | | | | | | | |
Db      271 IQRTLAAKSVSEAQKGIVFAAFLKLIVPFLVVLPGIAAYVITSDPQLMASLGDIATNLP 330

QY      315 ---EADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMEFARNIYQLSFRQN 371
      || | : |:: || : |::|::|| | : | ::| :|| :
Db      331 SAANADKAYPWLTQFL-PVGVGKGVFAALAAAI VSSLASMLNSTATIFTMDIYKEYISPD 389

QY      372 ASDKEIVVWMRITVFVFGASATAMALLTKTVYG--LWYLSSDLVYI-----VIFPQLLCV 424
      : | ::| | | : | :||: : | : || : : | : | |
Db      390 SGDHLKLVNV-----GRTAAVVALIIACLIAPMLGGIGQAFQYIQEYTGTVSPGILAV 441

QY      425 ----LFVKGTTNTYGAVAGYVSGFLRLITGGEPEY-LYLQPLIFYPGYYP-DDNGIYNQKFP 478
      || | | : ||: | | : | : |:: | | | | : | |
Db      442 FLLGLFWKKTTSKGAIIGVVASI-----PFALFLK---FMPLSMPFMDQMLYTLTLLFT 490

QY      479 FKTLAMVTSFLTNI 492
      : | || | : |
Db      491 MVVIAF-TSLSTSI 503

```

RESULT 10

SL52\_HUMAN

ID SL52\_HUMAN STANDARD; PRT; 672 AA.

AC P31639;

DT 01-JUL-1993 (Rel. 26, Created)

DT 01-JUL-1993 (Rel. 26, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Sodium/glucose cotransporter 2 (Na(+)/glucose cotransporter 2)

DE (Low affinity sodium-glucose cotransporter).

GN SLC5A2 OR SGLT2.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Kidney;

RX MEDLINE=93035768; PubMed=1415574;

RA Wells R.G., Pajor A.M., Kanai Y., Turk E., Wright E.M., Hediger M.A.;

RT "Cloning of a human kidney cDNA with similarity to the sodium-glucose

RT cotransporter.";

RL Am. J. Physiol. 263:F459-F465(1992).

CC -!- FUNCTION: Sodium-dependent glucose transporter. Has a Na+ to  
CC glucose coupling ratio of 1:1.

CC -!- FUNCTION: Efficient substrate transport in mammalian kidney is  
CC provided by the concerted action of a low affinity high capacity  
CC and a high affinity low capacity Na(+)/glucose cotransporter  
CC arranged in series along kidney proximal tubules.

CC -!- SUBCELLULAR LOCATION: Integral membrane protein.

CC -!- SIMILARITY: BELONGS TO THE SODIUM:SOLUTE SYMPORTER FAMILY (SSF).

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
 DR EMBL; M95549; AAA36608.1; -.  
 DR PIR; A56765; A56765.  
 DR Genew; HGNC:11037; SLC5A2.  
 DR MIM; 182381; -.  
 DR GO; GO:0016021; C:integral to membrane; TAS.  
 DR GO; GO:0005362; F:low-affinity glucose:sodium symporter activity; TAS.  
 DR GO; GO:0005975; P:carbohydrate metabolism; TAS.  
 DR GO; GO:0006810; P:transport; TAS.  
 DR InterPro; IPR001734; Na/solut\_symport.  
 DR Pfam; PF00474; SSF; 1.  
 DR TIGRFAMs; TIGR00813; sss; 1.  
 DR PROSITE; PS00456; NA\_SOLUT\_SYMP\_1; 1.  
 DR PROSITE; PS00457; NA\_SOLUT\_SYMP\_2; 1.  
 DR PROSITE; PS50283; NA\_SOLUT\_SYMP\_3; 1.  
 KW Transport; Sugar transport; Transmembrane; Sodium transport; Symport;  
 KW Glycoprotein.

FT	DOMAIN	1	25	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	26	44	POTENTIAL.
FT	DOMAIN	45	61	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	62	82	POTENTIAL.
FT	DOMAIN	83	102	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	103	123	POTENTIAL.
FT	DOMAIN	124	168	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	169	188	POTENTIAL.
FT	DOMAIN	189	205	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	206	226	POTENTIAL.
FT	DOMAIN	227	270	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	271	291	POTENTIAL.
FT	DOMAIN	292	314	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	315	334	POTENTIAL.
FT	DOMAIN	335	423	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	424	443	POTENTIAL.
FT	DOMAIN	444	455	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	456	476	POTENTIAL.
FT	DOMAIN	477	526	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	527	547	POTENTIAL.
FT	DOMAIN	548	650	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	651	671	POTENTIAL.
FT	CARBOHYD	250	250	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	399	399	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	SITE	40	40	IMPLICATED IN SODIUM COUPLING
FT				(BY SIMILARITY).
FT	SITE	300	300	IMPLICATED IN SODIUM COUPLING
FT				(BY SIMILARITY).
SQ	SEQUENCE	672 AA;	72896 MW;	233C65F1601B0337 CRC64;

Query Match 9.8%; Score 292; DB 1; Length 672;  
 Best Local Similarity 24.1%; Pred. No. 3.9e-13;  
 Matches 147; Conservative 91; Mismatches 237; Indels 136; Gaps 22;

QY 8 LIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGY 67

```

      :: | ::||:: ||:: | | | | : : || : | :: | : | | :
Db      26 ILVIAAYFLLVIGVGLWSMCRN-NGRTV----GGYFLAGRSMVWVPVGASLFASNIGSGH 80

Qy      68 INGTAEEAVYVPGYGLAWAQAPIGYSL-----LILGGLFFAKPMRSKGYVTMLDPFQQIYG 123
      | | | | | | | | | | : : : : | | | | : | : | | |
Db      81 FVGLA-----GTGAASGLAVAGFEWNALFVLLLGWLFAPVYLTAGVITM----PQYLR 130

Qy      124 KRMGG-----LLFIPALMGEMFWAAAF--SALGATISVIIDVDMHISVIISA 169
      | | | | | | | | | | : : : : | | | | : | | |
Db      131 KREGGRRIRLYLSVLSLFLYIFTKISVDMFSGAVFIQQALGWNI-----YASVIAL 182

Qy      170 LIATLYTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKY----- 224
      | :||: ||| :: ||| || | | | : : | : : | |
Db      183 GITMIYTVTGGLAALMYTDTVQTFVILGGACILMGYAFHEVG----GYSGLFDKYLGAAT 238

Qy      225 -----QKPWLGTVDSSSEVYSWLDSFLLL---MLGGIPW-----QAYF 258
      : | : | : | | | : | | : | : | |
Db      239 SLTVSEDPVGNISSEFCYRPRPDSYHLLRHPVTGDLPPWALLLGLTIVSGWYWCSDQVIV 298

Qy      259 QRVLSSSSATYAQVLSFLAAGCLVMAIPAILIGAIGASTDWNQTAYGLPDPKT----TE 314
      || | : | | : : | : | : : | | : : | : | |
Db      299 QRCLAGKSLTHIKAGCILCGYLKLTMPFLMVMPGMISRLYPDEVACVVPEVCRRVCSTE 358

Qy      315 E--ADMILPIVLQYLCPVYISFFGLGAVSAVMSSADSSILSASSMFARNIYQLSFRQNA 372
      :: : | : : | | : | : ||: || | | : | : | : | |
Db      359 VGCSNIAYPRLVVKLMPNGLRGLMLAVMLAALMSSLASIFNSSSTLFTMDIY-TRLRRA 417

Qy      373 SDKEIVWMRI-TVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQLLCV----LFV 427
      | :||: | | : ||: | : : : | : | : | : | |
Db      418 GDRELLLVGRLWVVFIVVSVAVLPPVQAAQGGQLFDYIQAVSSYLAPPVSAVFLALFV 477

Qy      428 KGTNTYGAAGYVSGFLFRITGGEPYLYLQPLIFYPGYPPDDNGIYNQKFPFKTLAMV-- 485
      | || | : || : : | : : : | | : : |
Db      478 PRVNEQGAFWGLIGLLMGLARLIP-----EFSFGSGSCVQP 514

Qy      486 ---TSFLTNICISYLAKYLFE-SGTLPKLDVFDVAV-----ARHSEENMDKTI 530
      : || : | | || || | : : | : | | : |
Db      515 SACPAFLCGVHYLYFAIVLFFCSGLLTTLTVSLCTAPIPRKHLHRLVFSRLHSKE----- 568

Qy      531 LVKNENIKLDE 541
      : | : | |
Db      569 --EREDLDADE 577

```

RESULT 11

SL54\_MOUSE

ID SL54\_MOUSE STANDARD; PRT; 656 AA.

AC Q9ET37;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Low affinity sodium-glucose cotransporter (Sodium/glucose

DE cotransporter 3) (Na+)/glucose cotransporter 3).

GN SLC5A4A.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=129/Sv;  
 RX MEDLINE=20499361; PubMed=11042146;  
 RA Pletcher M.T., Roe B.A., Chen F., Do T., Do A., Malaj E., Reeves R.H.;  
 RT "Chromosome evolution: the junction of mammalian chromosomes in the  
 RT formation of mouse chromosome 10.";  
 RL Genome Res. 10:1463-1467(2000).  
 CC -!- FUNCTION: Sodium-dependent glucose transporter (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -!- SIMILARITY: BELONGS TO THE SODIUM:SOLUTE SYMPORTER FAMILY (SSF).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF251267; AAG01741.1; -.  
 DR MGD; MGI:1927848; Slc5a4a.  
 DR InterPro; IPR001734; Na/solut\_symport.  
 DR Pfam; PF00474; SSF; 1.  
 DR TIGRFAMs; TIGR00813; sss; 1.  
 DR PROSITE; PS00456; NA\_SOLUT\_SYMP\_1; FALSE\_NEG.  
 DR PROSITE; PS00457; NA\_SOLUT\_SYMP\_2; 1.  
 DR PROSITE; PS50283; NA\_SOLUT\_SYMP\_3; 1.  
 KW Transport; Sugar transport; Transmembrane; Sodium transport; Symport;  
 KW Glycoprotein.  
 FT DOMAIN 1 28 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 29 47 POTENTIAL.  
 FT DOMAIN 48 64 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 65 85 POTENTIAL.  
 FT DOMAIN 86 105 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 106 126 POTENTIAL.  
 FT DOMAIN 127 171 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 172 191 POTENTIAL.  
 FT DOMAIN 192 208 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 209 229 POTENTIAL.  
 FT DOMAIN 230 270 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 271 291 POTENTIAL.  
 FT DOMAIN 292 314 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 315 334 POTENTIAL.  
 FT DOMAIN 335 423 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 424 443 POTENTIAL.  
 FT DOMAIN 444 455 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 456 476 POTENTIAL.  
 FT DOMAIN 477 526 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 527 547 POTENTIAL.  
 FT DOMAIN 548 634 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 635 655 POTENTIAL.  
 FT CARBOHYD 248 248 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 656 AA; 71837 MW; A6668E815204D39B CRC64;

Query Match

9.8%; Score 290; DB 1; Length 656;

Best Local Similarity 22.3%; Pred. No. 5.2e-13;  
Matches 147; Conservative 102; Mismatches 252; Indels 158; Gaps 25;

```

Qy      11 IIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGYING 70
      |:::|||||:| | : || : |::|:|:| |
Db      32 IIVIYFVVVMAVGWVWAMLKTNRSTVG-----GFFLAGRSMTWWPMGASLFASNIGSGHFGV 86

Qy      71 TAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSK-GYVTMLDPFQQIYGKRMGG- 128
      | : | :| |::| |::| | : | |::| : || ||
Db      87 LAGTGAASGIAVT-AFESHSEFALLLVLGWIFV--PIYIKAGVMTM----PEYLKKRFGGK 139

Qy     129 -----LLFIPALM---GEMFWAAAIIFSALGATISVIIDVDMHISVIISALIATLYT 176
      ||| :: :| : ||| | : : :::::| | ::|
Db     140 RLQIYLSILFLFICVILTISADIF-SGAIF-----IKLALGLNLYLAILILLAITAIFT 192

Qy     177 LVGGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVD--- 233
      : ||| || ||| | : || :| : || | : : | : | :
Db     193 ITGGLASVIYTDTVQAVIMLVGSFILMVFAF----VEVGGYESFTEKFMNAIPSVVEGDN 248

Qy     234 ---SSEVYS-WLDSFLLL---MLGGIPW-----QAYFQRLVSSSSAT 268
      :| | : ||| : : | ||| | ||| : :
Db     249 LTINSRCYTPQPDSEHIFRDPVTGDIWPFGTAFGMPITALWYWCINQVIVQRCICGKNLS 308

Qy     269 YAQVLSFLAAFGLVMAIPAILIGAIGASTDWNQTAYGLP-----DPKTTEEADMI 319
      : : | : | : :| | : | :| | | |
Db     309 HVKAACILCGYLKLLPLFFMVMPGMISRILYTDVMVACVVPSECVKHCGVDVGCTNYA--- 365

Qy     320 LPIVLQYLCVPYISFFGLGAVSAAVMSSADSSILSASSMFARNIYQLSFRQNASDKEIVW 379
      |::: | | : | : |::||| | |||::| ::| | : ||::|::
Db     366 YPMLVLKLMPPGLRGLMLSVMLASLMSSLTSVFNSASTLFTIDLY-TKIRKKASERELLI 424

Qy     380 VMRITVFVFGASATAMALLTKTVYG---LWYLSSDLVYI--VIFPQLLCVLFVKGTNTYG 434
      | : | | : : : : | : | : | : | : | | | |
Db     425 AGRLFVSVLIVTSILWVPIVEVSQGGQLVHYTEAISSYLGPPIAAVFLVAVFCKRANEQG 484

Qy     435 AVAGYVSGFLF-----RITGGEPYLYLQPLIFYPGYYPDDNG 471
      | | : || : :| | ||| ::|:
Db     485 AFWGLMVGLVMGLIRMIAEFSYGTGSC LAPSSCPKIICGVHYLYFAIILFF----- 535

Qy     472 IYNQKFPPFKTLAMVTSFLTNICISYLAKYLFESGTLPPKLDV----FDAVVARHSEENMD 527
      | : :||| | | || : || :|
Db     536 -----VCILVILGVSYLTK-----PIPDVHLHRLCWALRNSKEERID 572

Qy     528 KTILVKNNEN---IKLDELALVKPR-----QSMTLSSTFTNKEAFLDVDSSP 570
      | || : :| ||| | | :|| | |
Db     573 LDAEDKEENGADDRTEEDQTEKPRGCLKKTCDLFCGLQRAEFKLTKEVEEALDTEKTP 631

```

# RESULT 12

SL53\_MOUSE

ID SL53\_MOUSE STANDARD; PRT; 718 AA.

AC Q9JKZ2;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Sodium/myo-inositol cotransporter (Na+)/myo-inositol cotransporter).

GN SLC5A3.

OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20237552; PubMed=10773690;  
 RA McVeigh K.E., Mallee J.J., Lucente A., Barnoski B.L., Wu S.,  
 RA Berry G.T.;  
 RT "Murine chromosome 16 telomeric region, homologous with human  
 RT chromosome 21q22, contains the osmoregulatory Na(+)/myo-inositol  
 RT cotransporter (SLC5A3) gene.";  
 RL Cytogenet. Cell Genet. 88:153-158(2000).  
 CC -!- FUNCTION: Prevents intracellular accumulation of high  
 CC concentrations of myo-inositol (an osmolyte) that result in  
 CC impairment of cellular function.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -!- SIMILARITY: BELONGS TO THE SODIUM:SOLUTE SYMPORTER FAMILY (SSF).  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 DR EMBL; AF220915; AAF43668.1; -.  
 DR MGD; MGI:1858226; Slc5a3.  
 DR InterPro; IPR001734; Na/solut\_sympor.  
 DR Pfam; PF00474; SSF; 1.  
 DR TIGRFAMs; TIGR00813; sss; 1.  
 DR PROSITE; PS00456; NA\_SOLUT\_SYMP\_1; 1.  
 DR PROSITE; PS00457; NA\_SOLUT\_SYMP\_2; 1.  
 DR PROSITE; PS50283; NA\_SOLUT\_SYMP\_3; 1.  
 KW Transport; Transmembrane; Sodium transport; Symport; Glycoprotein.  
 FT DOMAIN 1 9 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 10 29 POTENTIAL.  
 FT DOMAIN 30 38 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 39 57 POTENTIAL.  
 FT DOMAIN 58 86 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 87 110 POTENTIAL.  
 FT DOMAIN 111 123 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 124 144 POTENTIAL.  
 FT DOMAIN 145 157 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 158 183 POTENTIAL.  
 FT DOMAIN 184 186 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 187 205 POTENTIAL.  
 FT DOMAIN 206 303 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 304 324 POTENTIAL.  
 FT DOMAIN 325 353 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 354 376 POTENTIAL.  
 FT DOMAIN 377 406 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 407 430 POTENTIAL.  
 FT DOMAIN 431 443 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 444 462 POTENTIAL.  
 FT DOMAIN 463 510 CYTOPLASMIC (POTENTIAL).



FT	TRANSMEM	511	532	POTENTIAL.
FT	DOMAIN	533	695	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	696	716	POTENTIAL.
FT	CARBOHYD	32	32	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	SITE	24	24	IMPLICATED IN SODIUM COUPLING
FT				(BY SIMILARITY).
FT	SITE	285	285	IMPLICATED IN SODIUM COUPLING
FT				(BY SIMILARITY).
SQ	SEQUENCE	718 AA;	79554 MW;	D035CFBECDDA803B CRC64;

Query Match 9.7%; Score 289; DB 1; Length 718;  
 Best Local Similarity 21.7%; Pred. No. 6.8e-13;  
 Matches 150; Conservative 113; Mismatches 209; Indels 218; Gaps 32;

Qy	9	IAII-VFYLLILLVGIWAAWRKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGG-	66
		: : : :   : :   :   : :   :   :   :   :	
Db	10	IADVVALYFILVMCIGFFAMWKSNRSTVS-----GYFLAGRSM--TWVAIGA	53
Qy	67	--YINGTAEAVYVPGYGLAWAQAPIGYS-----LSLILGGLFFAKPMRSKGYVTM	114
		: : : : : :   :   :   : :   :   :   :   :	
Db	54	SLFVSNIGSEHFI---GLAGSGAASGFAVGAWEFNALLLQLLGWVFIPYIYIRS-GVYTM	109
Qy	115	LDPFQQIYGKRMGG-----LLFIPALMGEMFWAAAFSALGATISVIIDVDMH	162
		:         : : : :   : : : :   : : : :	
Db	110	----PEYLSKRFGGHRIQVYFAALSLLLYIFTKLSVDLYSGALF-----IQESLGWNLY	159
Qy	163	ISVIISALIATLYTLVGGLYSVAYTDVVQLFCIFVG---LWISV-----PFAL	207
		:     : :   :   :   :   :   :   :   :	
Db	160	VSVILLIGMTALLTVTGGLVAVIYTDTLQALLMIIGALTLMVISMVKIGGFEEVKRRYML	219
Qy	208	SHPAVADI-----GFTAVHAKYQK-----PWLGTV---DSSEVYSWLD	243
		:           :   :   :   :   :   :	
Db	220	ASPDVASILLKYNLSNTNACMVHPKANALKMLRDPDDEDVPWPGFILGQTPASVWYWCAD	279
Qy	244	FLLLMLGGIPWQAYFQRVLSSSSATYQ---VLSFLAAFGCLVMAIPAIL-----	290
		: : : :   : : : :   : : : :	
Db	280	-----QVIVQRVLAAKNIAHAKGSTLMAGFLKLLPMFIIVVPGMISRIVFADEI	328
Qy	291	-----IGAIGASTDWNQTAYGLPDKTTEEADMILPIVLQYLCVPYISFFGLGAVSA	342
		:   : :   :   :   :   :   :   :	
Db	329	ACINPEHCMQVCGSRAGCSNIAY-----PRLVMTLVFVGLRGLMMAVMIA	373
Qy	343	AVMSSADSSILSASSMFARNIYQLSFRQNASDKEIVVMRITV-FVFGASATAMALLTKT	401
		:         : : : :   : : :   : : :   : : :	
Db	374	ALMSDLDISIFNSASTIFTLDVYKL-IRKSASSRELMIVGRIFVAFMVVISIAWVPIIVEM	432
Qy	402	VYGLWYLSSDLVYIVIFPQL----LCVLFVKGTNT----YGAVAGYVSG---LFLRITGG	450
		: : : :   :       :   :   :   :	
Db	433	QGGQMYLYIQEVADYLTTPVAALFLLAIFWKRCNEQGAFYGGMAGFVLGAVRLILAFTYR	492
Qy	451	EP-----YLYLQPLIFYPGYPPDDNGIYNQKFPFKTLAMVTSFLTNICI	494
		: : : :   : : : :   : : : :	
Db	493	APECDQPDNRPGFIKDIHYMYVATALFW-----ITGLIT-VIV	529
Qy	495	SYLAKYLFESGTLPPKLDVFDVAVVARHSEENMDKTILV-----KNENIKLDELALVK-P	547
		: : : :   :   :   : : :	
Db	530	SLL-----TPPPTKDQI-----RTTTFWSKTLVTKESCSQKDEPYKMQEKSILQCS	576

Qy 548 RQSM T L S S T F T N K E A F L D V D S S P E G S G T E D 577  
 | : | | | : : : | : | |  
 Db 577 E N S E V I S H T I P N G K S --- E D S I K G L Q P E D 602

RESULT 13

OPUE\_BACSU

ID OPUE\_BACSU STANDARD; PRT; 492 AA.  
 AC O06493;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Osmoregulated proline transporter (Sodium/proline symporter).  
 GN OPUE OR BSU06660.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168 / JH642;  
 RA von Blohn C., Kempf B., Kappes R.M., Bremer E.;  
 RT "Osmotress response in Bacillus subtilis: characterization of a  
 RT proline uptake system (OpuE) regulated by high osmolarity and the  
 RT alternative transcription factor sigma B.";  
 RL Mol. Microbiol. 25:175-187(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=97124186; PubMed=8969499;  
 RA Borriss R., Porwollik S., Schroeter R.;  
 RT "The 52 degrees-55 degrees segment of the Bacillus subtilis  
 RT chromosome: a region devoted to purine uptake and metabolism, and  
 RT containing the genes cotA, gabP and guaA and the pur gene cluster  
 RT within a 34960 bp nucleotide sequence.";  
 RL Microbiology 142:3027-3031(1996).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=98044033; PubMed=9384377;  
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
 RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
 RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,  
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,  
 RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
 RA Guiseppe G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
 RA Kobayashi Y., Koetter P., Konigstein G., Krogh S., Kumano M.,  
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,

RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
 RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,  
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,  
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;  
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*  
 RT *subtilis*.";  
 RL Nature 390:249-256(1997).

CC -!- FUNCTION: CATALYZES THE SODIUM-DEPENDENT UPTAKE OF EXTRACELLULAR  
 CC PROLINE.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -!- SIMILARITY: BELONGS TO THE SODIUM:SOLUTE SYMPORTER FAMILY (SSF).

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; U92466; AAB66512.1; -.  
 DR EMBL; AF011545; AAB72182.1; -.  
 DR EMBL; Z99107; CAB12486.1; -.  
 DR PIR; H69670; H69670.  
 DR SubtiList; BG12641; opuE.  
 DR InterPro; IPR001734; Na/solut\_symport.  
 DR Pfam; PF00474; SSF; 1.  
 DR TIGRFAMs; TIGR00813; sss; 1.  
 DR PROSITE; PS00456; NA\_SOLUT\_SYMP\_1; 1.  
 DR PROSITE; PS00457; NA\_SOLUT\_SYMP\_2; 1.  
 DR PROSITE; PS50283; NA\_SOLUT\_SYMP\_3; 1.  
 KW Transport; Transmembrane; Sodium transport; Symport;  
 KW Complete proteome.

FT	TRANSMEM	3	23	POTENTIAL.
FT	TRANSMEM	62	82	POTENTIAL.
FT	TRANSMEM	125	145	POTENTIAL.
FT	TRANSMEM	161	181	POTENTIAL.
FT	TRANSMEM	191	211	POTENTIAL.
FT	TRANSMEM	224	244	POTENTIAL.
FT	TRANSMEM	271	291	POTENTIAL.
FT	TRANSMEM	314	334	POTENTIAL.
FT	TRANSMEM	365	385	POTENTIAL.
FT	TRANSMEM	394	414	POTENTIAL.
FT	TRANSMEM	424	444	POTENTIAL.
FT	TRANSMEM	449	469	POTENTIAL.
SQ	SEQUENCE	492 AA;	53282 MW;	23459873F1E799E6 CRC64;

Query Match 9.6%; Score 285; DB 1; Length 492;  
 Best Local Similarity 22.1%; Pred. No. 8.5e-13;  
 Matches 118; Conservative 97; Mismatches 214; Indels 106; Gaps 18;



RT "Cloning of the cDNA for a Na<sup>+</sup>/myo-inositol cotransporter, a  
 RT hypertonicity stress protein.";  
 RL J. Biol. Chem. 267:6297-6301(1992).  
 CC -!- FUNCTION: Prevents intracellular accumulation of high  
 CC concentrations of myo-inositol (an osmolyte) that result in  
 CC impairment of cellular function.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -!- TISSUE SPECIFICITY: Brain and kidney.  
 CC -!- INDUCTION: Medium hypertonicity.  
 CC -!- SIMILARITY: BELONGS TO THE SODIUM:SOLUTE SYMPORTER FAMILY (SSF).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M85068; -; NOT\_ANNOTATED\_CDS.  
 DR PIR; A42163; A42163.  
 DR InterPro; IPR001734; Na/solut\_sympoort.  
 DR Pfam; PF00474; SSF; 1.  
 DR TIGRFAMs; TIGR00813; sss; 1.  
 DR PROSITE; PS00456; NA\_SOLUT\_SYMP\_1; 1.  
 DR PROSITE; PS00457; NA\_SOLUT\_SYMP\_2; 1.  
 DR PROSITE; PS0283; NA\_SOLUT\_SYMP\_3; 1.  
 KW Transport; Transmembrane; Sodium transport; Symport; Glycoprotein.  
 FT DOMAIN 1 9 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 10 29 POTENTIAL.  
 FT DOMAIN 30 38 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 39 57 POTENTIAL.  
 FT DOMAIN 58 86 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 87 110 POTENTIAL.  
 FT DOMAIN 111 123 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 124 144 POTENTIAL.  
 FT DOMAIN 145 157 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 158 183 POTENTIAL.  
 FT DOMAIN 184 186 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 187 205 POTENTIAL.  
 FT DOMAIN 206 303 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 304 324 POTENTIAL.  
 FT DOMAIN 325 353 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 354 376 POTENTIAL.  
 FT DOMAIN 377 406 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 407 430 POTENTIAL.  
 FT DOMAIN 431 443 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 444 462 POTENTIAL.  
 FT DOMAIN 463 510 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 511 532 POTENTIAL.  
 FT DOMAIN 533 695 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 696 716 POTENTIAL.  
 FT CARBOHYD 32 32 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT SITE 24 24 IMPLICATED IN SODIUM COUPLING  
 FT (BY SIMILARITY).  
 FT SITE 285 285 IMPLICATED IN SODIUM COUPLING  
 FT (BY SIMILARITY).

Query Match 9.4%; Score 278.5; DB 1; Length 718;  
Best Local Similarity 23.1%; Pred. No. 3.7e-12;  
Matches 154; Conservative 115; Mismatches 222; Indels 177; Gaps 34;

RESULT 15  
SL53 BOVIN

ID SL53\_BOVIN STANDARD; PRT; 718 AA.  
AC P53793;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE Sodium/myo-inositol cotransporter (Na(+)/myo-inositol cotransporter).  
GN SLC5A3 OR SMIT.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Mallee J.J., Parrella T., Kwon H.M., Berry G.T.;  
RL Submitted (APR-1996) to the EMBL/GenBank/DDBJ databases.  
CC -!- FUNCTION: Prevents intracellular accumulation of high  
CC concentrations of myo-inositol (an osmolyte) that result in  
CC impairment of cellular function.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -!- SIMILARITY: BELONGS TO THE SODIUM:SOLUTE SYMPORTER FAMILY (SSF).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U41338; AAA93188.1; -.  
DR InterPro; IPR001734; Na/solut\_symport.  
DR Pfam; PF00474; SSF; 1.  
DR TIGRFAMs; TIGR00813; sss; 1.  
DR PROSITE; PS00456; NA\_SOLUT\_SYMP\_1; 1.  
DR PROSITE; PS00457; NA\_SOLUT\_SYMP\_2; 1.  
DR PROSITE; PS0283; NA\_SOLUT\_SYMP\_3; 1.  
KW Transport; Transmembrane; Sodium transport; Symport; Glycoprotein.  
FT DOMAIN 1 9 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 10 29 POTENTIAL.  
FT DOMAIN 30 38 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 39 57 POTENTIAL.  
FT DOMAIN 58 86 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 87 110 POTENTIAL.  
FT DOMAIN 111 123 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 124 144 POTENTIAL.  
FT DOMAIN 145 157 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 158 183 POTENTIAL.  
FT DOMAIN 184 186 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 187 205 POTENTIAL.  
FT DOMAIN 206 303 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 304 324 POTENTIAL.  
FT DOMAIN 325 353 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 354 376 POTENTIAL.  
FT DOMAIN 377 406 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 407 430 POTENTIAL.  
FT DOMAIN 431 443 EXTRACELLULAR (POTENTIAL).

FT	TRANSMEM	444	462	POTENTIAL.
FT	DOMAIN	463	510	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	511	532	POTENTIAL.
FT	DOMAIN	533	695	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	696	716	POTENTIAL.
FT	CARBOHYD	32	32	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	SITE	24	24	IMPLICATED IN SODIUM COUPLING
FT				(BY SIMILARITY).
FT	SITE	285	285	IMPLICATED IN SODIUM COUPLING
FT				(BY SIMILARITY).
SQ	SEQUENCE	718 AA; 79673 MW; 206BE25FA385111D CRC64;		

Query Match 9.3%; Score 275; DB 1; Length 718;  
 Best Local Similarity 22.2%; Pred. No. 6.6e-12;  
 Matches 148; Conservative 122; Mismatches 225; Indels 172; Gaps 32;

Qy	9	IAII-VFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGG-	66
		: : : :   : :   :   :   :   :   :   :   :   :   :   :	
Db	10	IAIVALYFILVMCIGFFAMWKSNRSTVS-----GYFLAGRSM--TWVAIGA	53
Qy	67	--YINGTAEAVYVPGYGLAWAQAPIGYS-----LSLILGGLFFAKPMRSKGYVTM	114
		: : : : :       :   :   :   :   :   :   :   :   :   :	
Db	54	SLFVSNIGSEHFT---GLAGSGAASGFAVGAWEFNALLLLQLLGWVFIPIYIRS-GVYTM	109
Qy	115	LDPFQQIYGKRMGG-----LLFIPALMGEMFWAAAIIFSALGATISVIIDVDMH	162
		:       :   :   : :   :   :   :   :   :   :   :   :	
Db	110	----PEYLSKRFGGHRIQVYFAALSILYIFTKLSVDLYSGALF-----IQESMGWNLY	159
Qy	163	ISVIISALIATLYTLVGGLYSVAYTDVVQLFCIFVG----LWISV-----PFAL	207
		:       :   :       :       :     :   :   :   :   :	
Db	160	VSVILLIGMTALLTVTGGLVAVIYDTRLQALLMIVGALTMVISMMEIGGFEEVKRRYML	219
Qy	208	SHPAVADIGFTAVHAKYQKPWLGTVDSSSEVYSWLDSFLLL-----MLGGIP-	253
		:           :   :   :   :   :   :   :   :   :   :	
Db	220	ASPNVTSILLT-----YN---LSNTNSCNVHPKKDALKMLRNPTDEDVPWPFGFILGQTPA	271
Qy	254	----W---QAYFQRVLSSSSATYAQ----VLSFLAAGCLVMAIPAIL-----	290
		: : : :   : :   : :   : :   : :   : :	
Db	272	SVWYWCADQVIVQRVLAANKIAHAKGSTLMAGFLKLLPMFIIVVPGMISRILFADDIACI	331
Qy	291	-----IGAIGASTDWNQTAYGLPDPKTTEEADMILPIVLQYLCPVYISFFGLGAVSAVM	345
		:   : :   :   :   :   :   :   :   :   :   :   :	
Db	332	NPEHCMQVCGSRAGCSNIAY-----PRLVMKLVVPVGLRGLMMAVMIAALM	376
Qy	346	SSADSSILSASSMFARNIYQLSFRQNASDKEIVWMRITV-FVFGASATAMALLTKTVYG	404
		: : :   : :   : :   : :   : :   : :   : :	
Db	377	SDLDSIFNSASTIFTLDVYKL-IRKSASSRELMIVGRI FVAFMVVISIAWVPIIEMQGG	435
Qy	405	LWYLSSDLVYIVIFPQL----LCVLFVKGTNTYGA VAGYVSGLFL---RITGGEPYLYLQ	457
		:   :   :   :   :   :   :   :   :   :   :	
Db	436	QMYLYIQEVADYLTTPVAALFLLAIFWKRCNEQGA FYGGMAGFILVVVRLT--LAFAYRA	493
Qy	458	PLIFYPGYPPDDNGIYNQKFPFKTLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDVAV	517
		: : : : : :   : :   : :   : :   : :   : :	
Db	494	P----ECDQPDNRPVFIKDIHYMYVATALFWITGL-ITVIVSLL---TPPPTKEQI---	541
Qy	518	VARHSEENMDKTILV-----KNENIKLDELALVK-PRQSM TLSSTFTNKEAFLDVDSSP	570



	:	::	:	:	:::	::	::	:
Db	542	--RTTTFWSKKSLVVKESCPKDEPYKMQEKSILRCSENSEVINHVIPNGKS----	EDSI	595				
Qy	571	EGSGTED	577					
		:						
Db	596	KGLQPED	602					

Search completed: March 22, 2004, 15:32:58  
 Job time : 32 secs